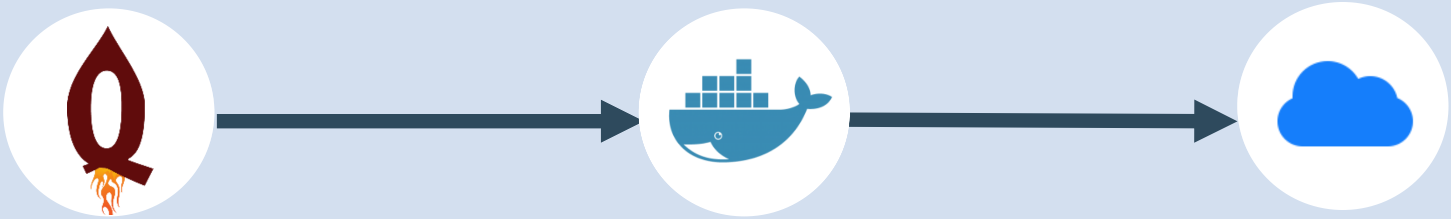
Building a Global-Scale Multi-Tenant Cloud Platform on AWS and Docker: Lessons Learned

Felix Gessert, Florian Bücklers

{fg,fb}@baqend.com

@baqendcom



Part One

Baqend & Our Infrastructure

Part Two

Docker Concepts

Part Three

Clustering: AWS ECS vs. Docker Swarm

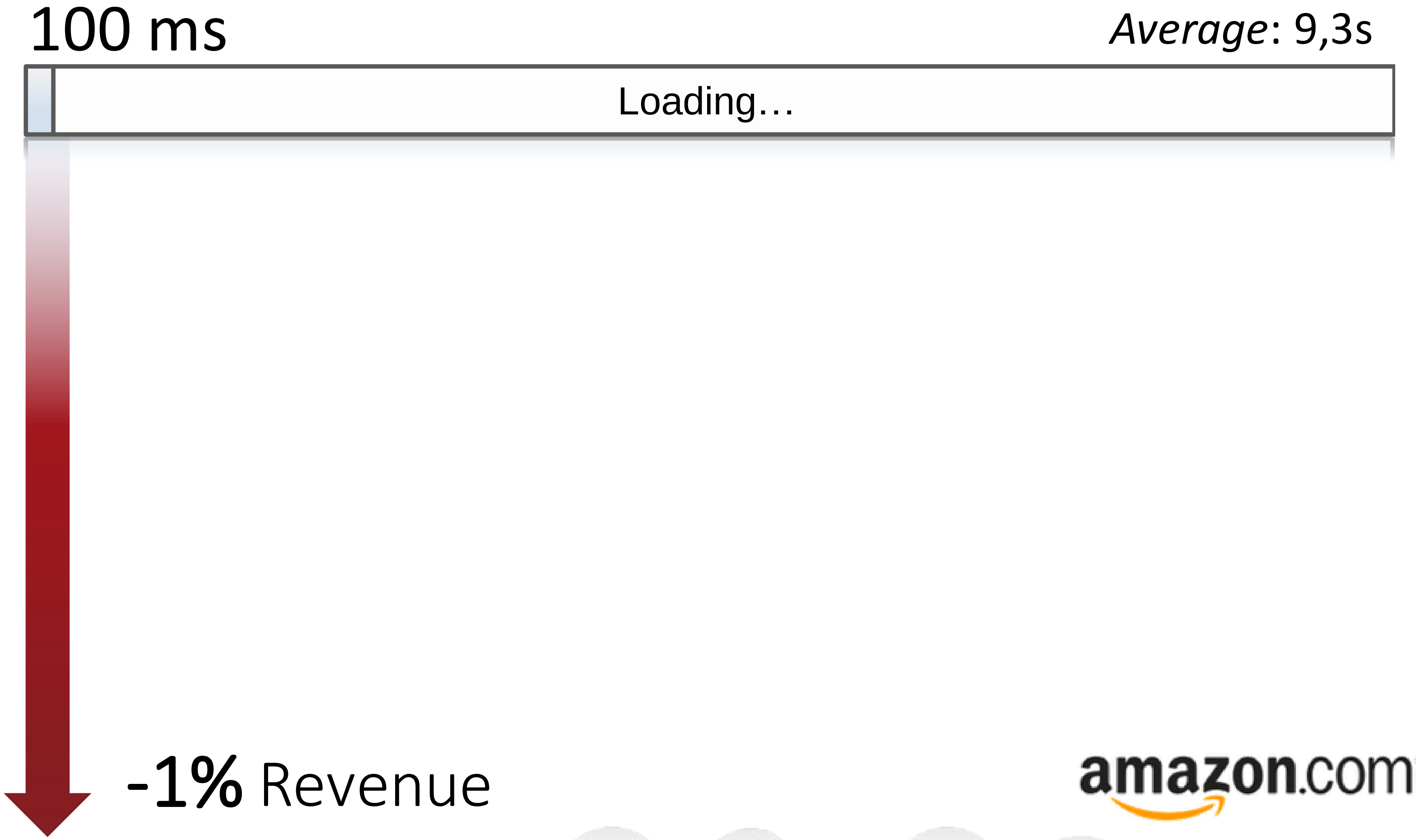
Presentation
is loading

The Latency Problem

Average: 9,3s

Loading…

The Latency Problem



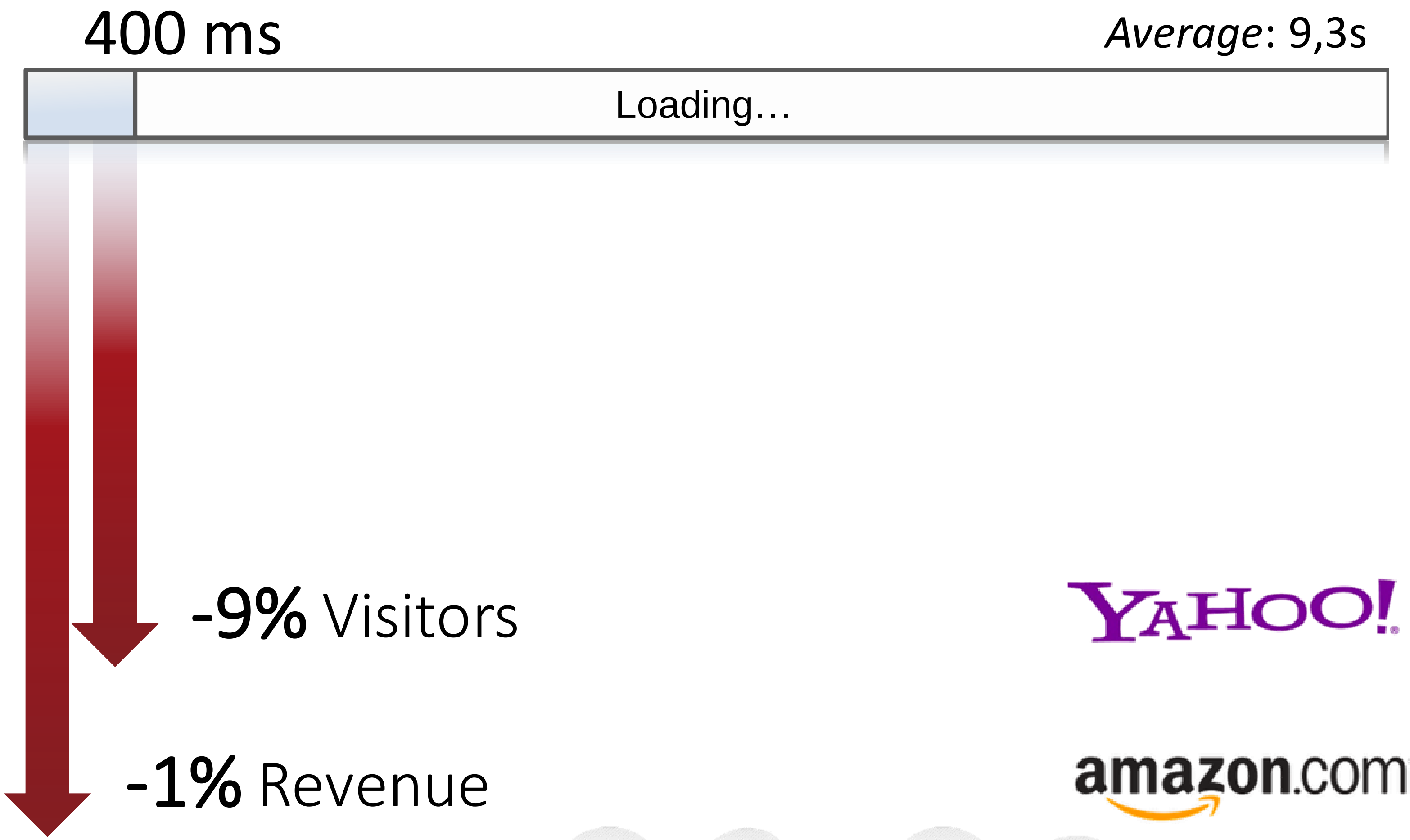
100 ms

Average: 9,3s

Loading…

-1% Revenue

The Latency Problem



400 ms

Average: 9,3s

Loading…

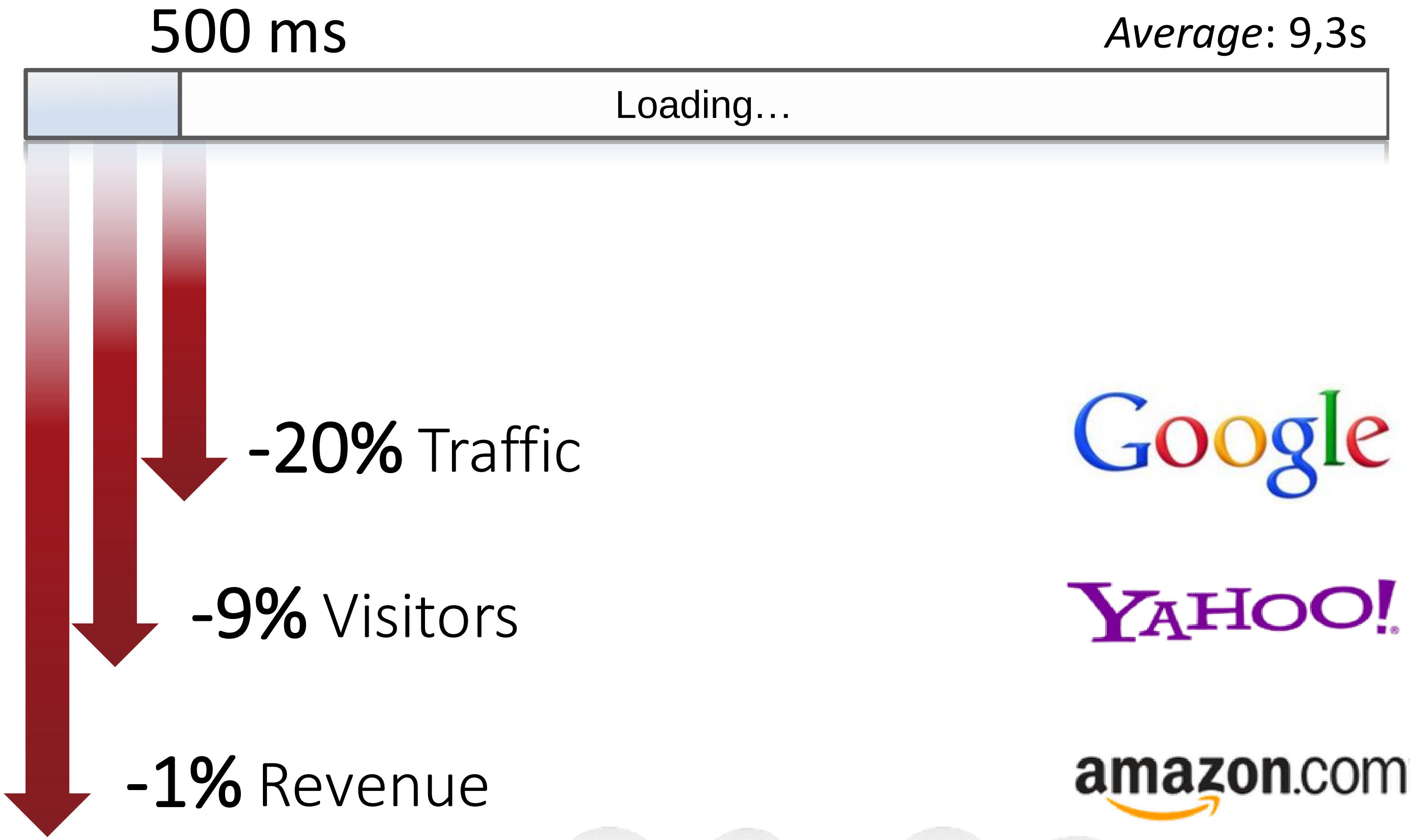
-9% Visitors

-1% Revenue

YAHOO!

amazon.com

The Latency Problem



500 ms

Average: 9,3s

Loading…

-20% Traffic

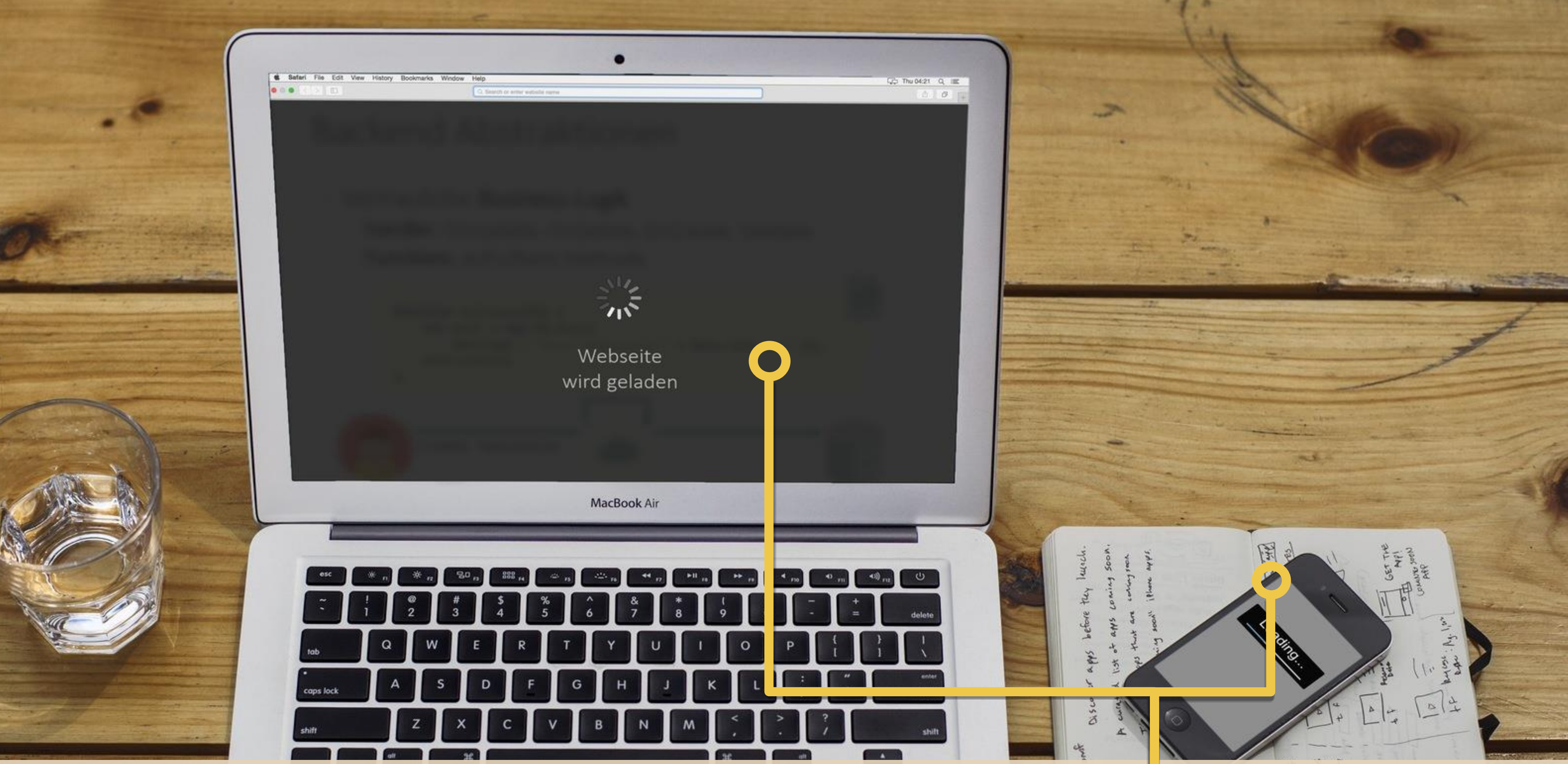
-9% Visitors

-1% Revenue

Google

YAHOO!

amazon.com

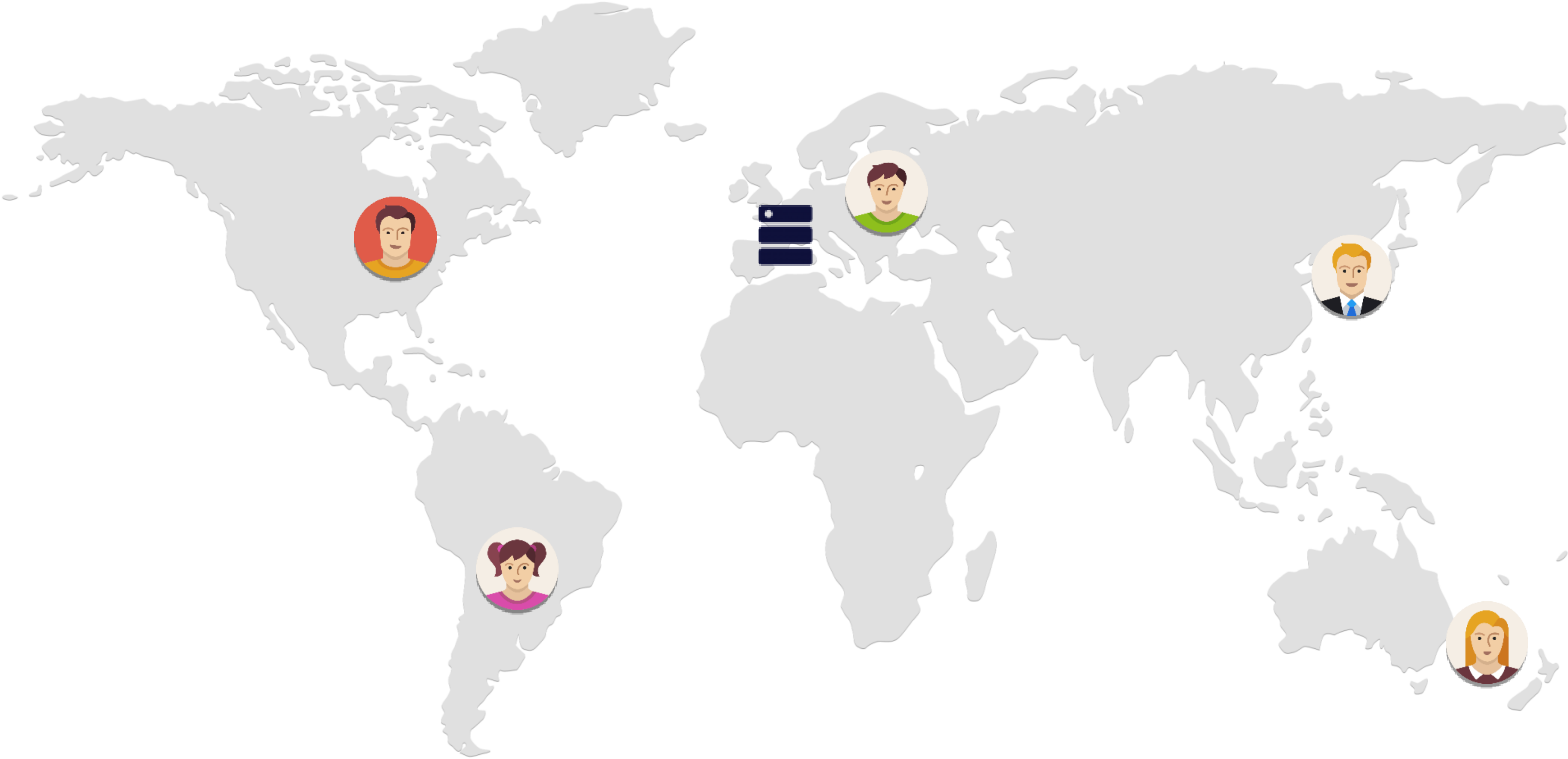


If perceived speed is such an import factor

...what causes slow page load times?

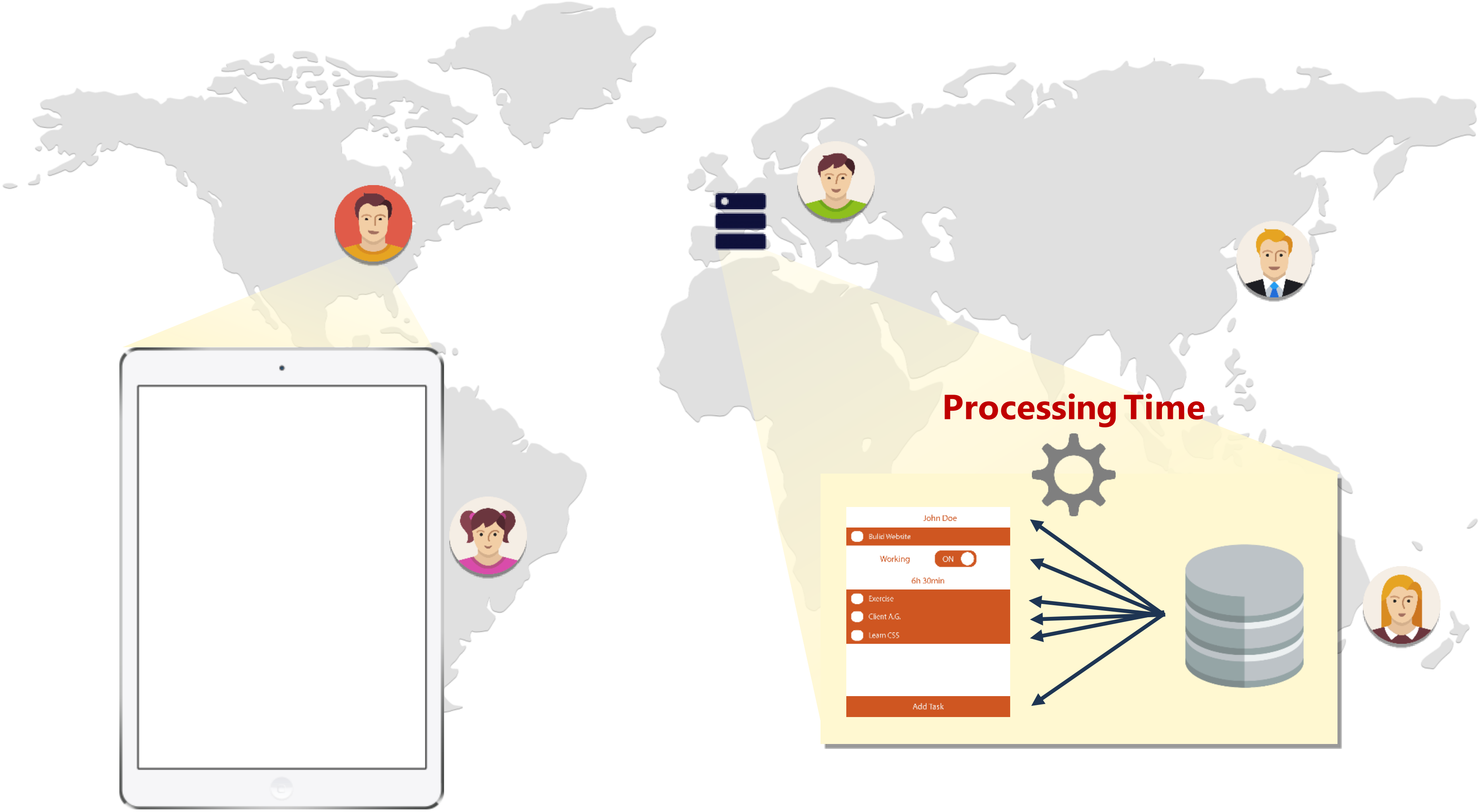
State of the Art

Two bottlenecks: latency und processing



State of the Art

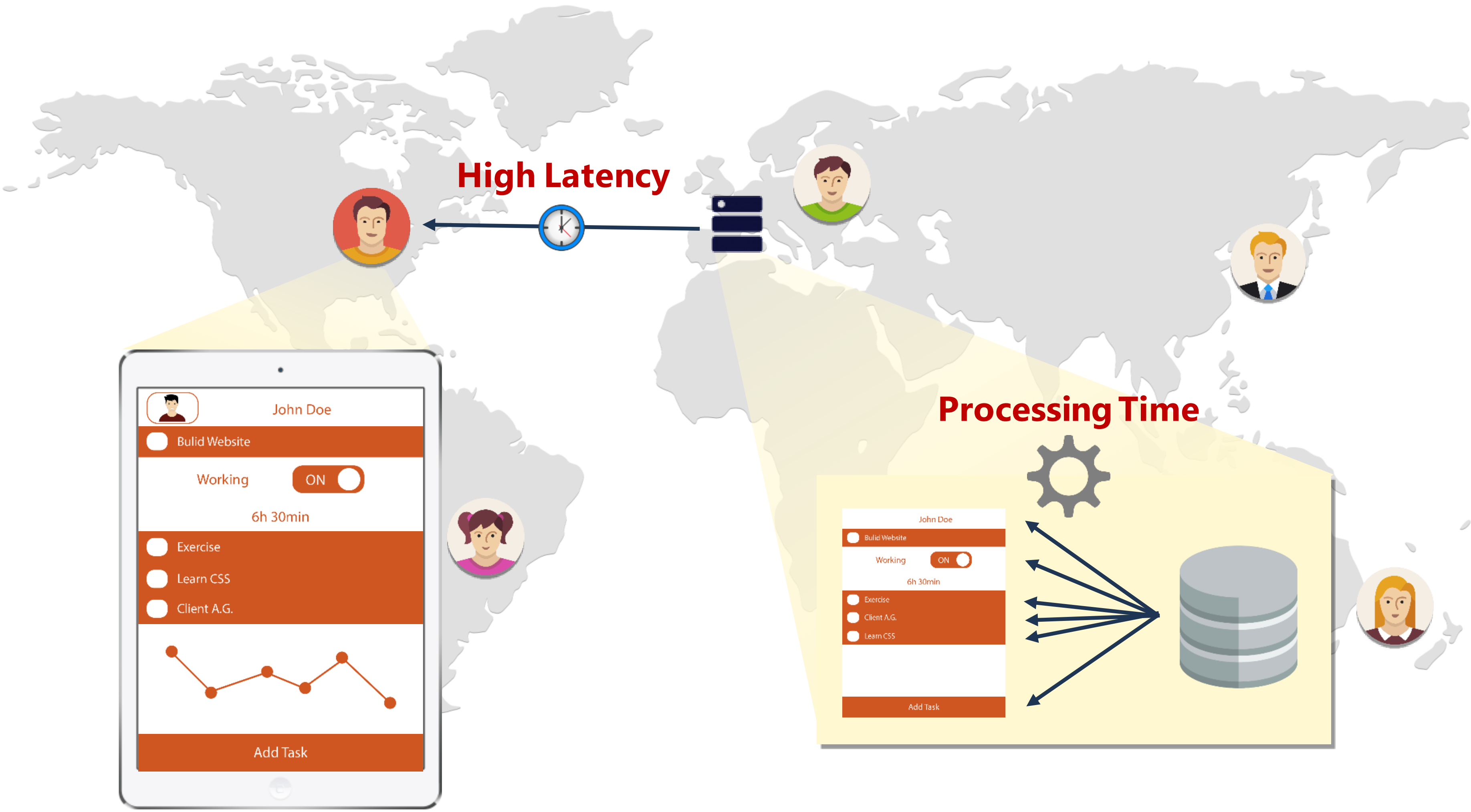
Two bottlenecks: latency und processing



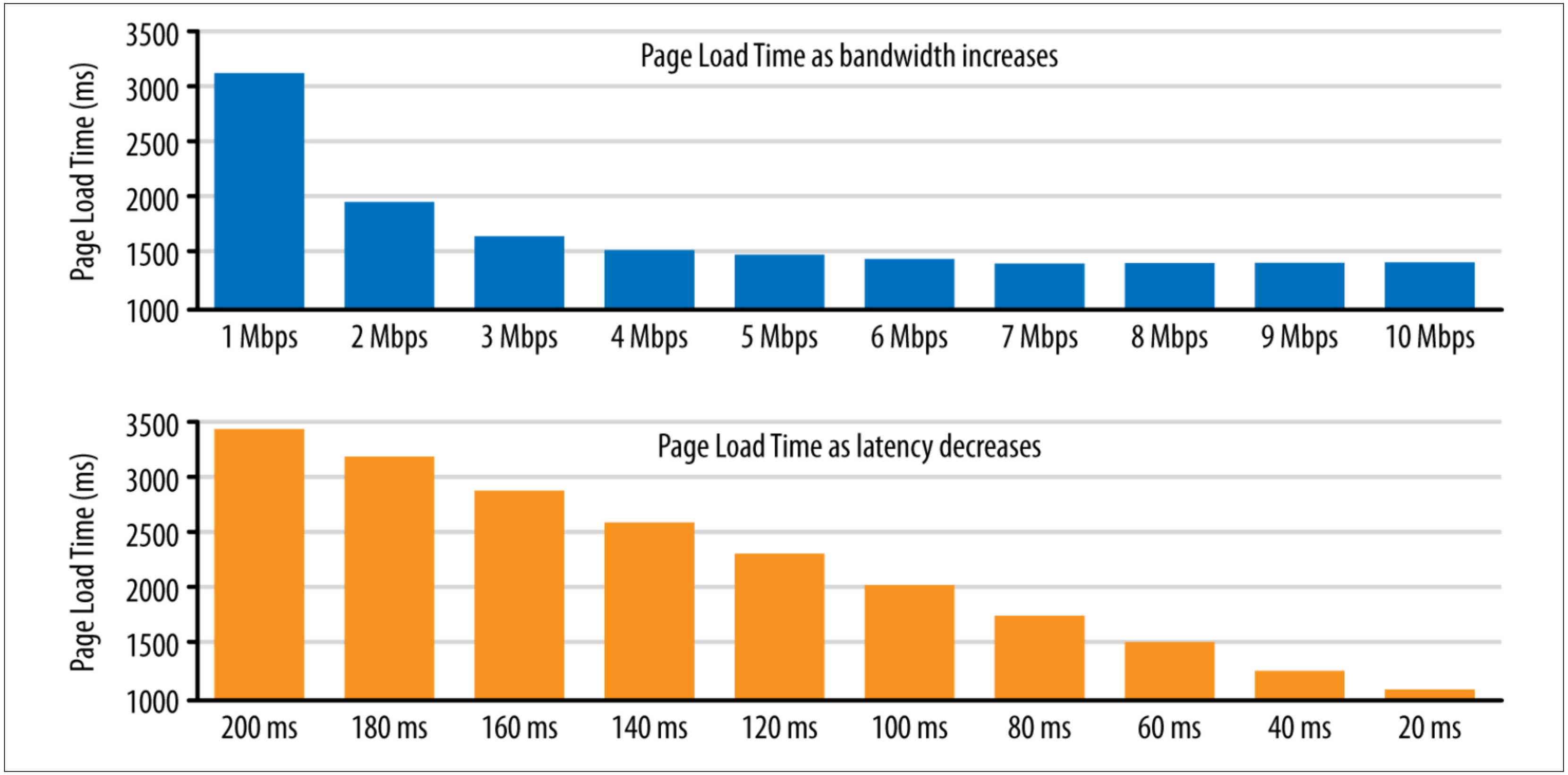
Processing Time

State of the Art

Two bottlenecks: latency und processing

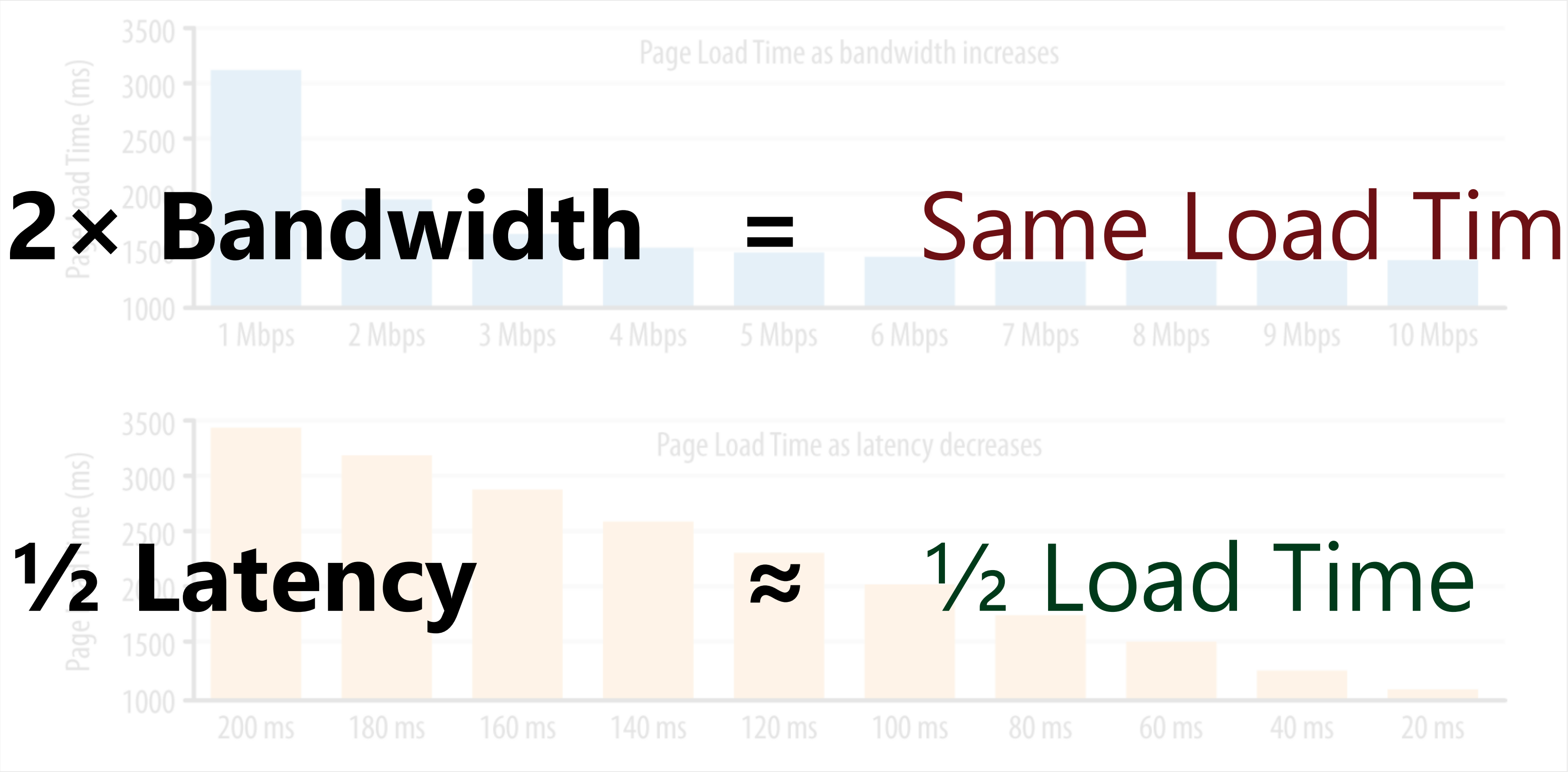


Problem: Network Latency



I. Grigorik, High performance browser networking. O'Reilly Media, 2013.

Problem: Netzwerklatenz



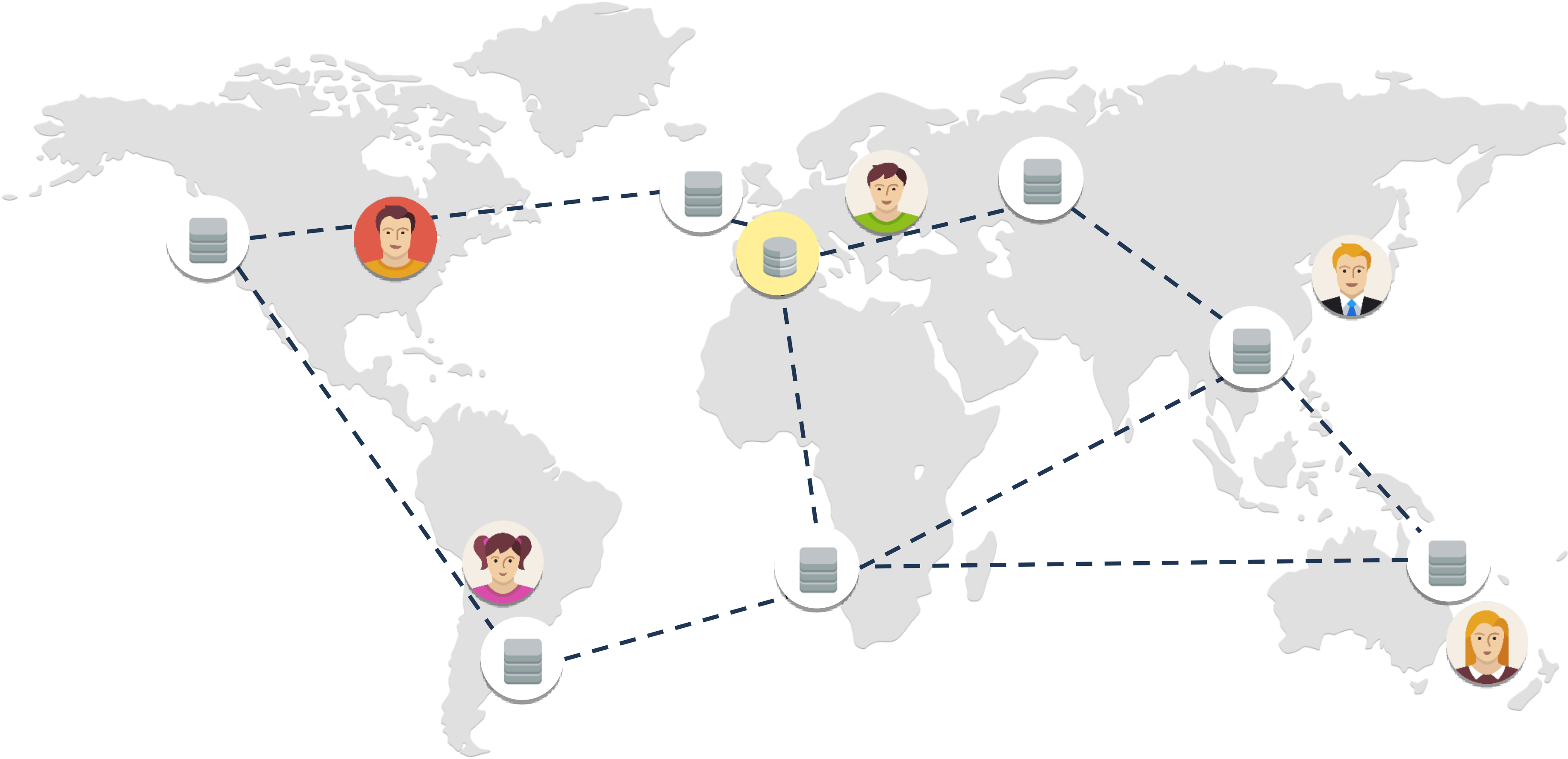
2× Bandwidth = Same Load Time

½ Latency ≈ ½ Load Time

I. Grigorik, High performance browser networking. O'Reilly Media, 2013.

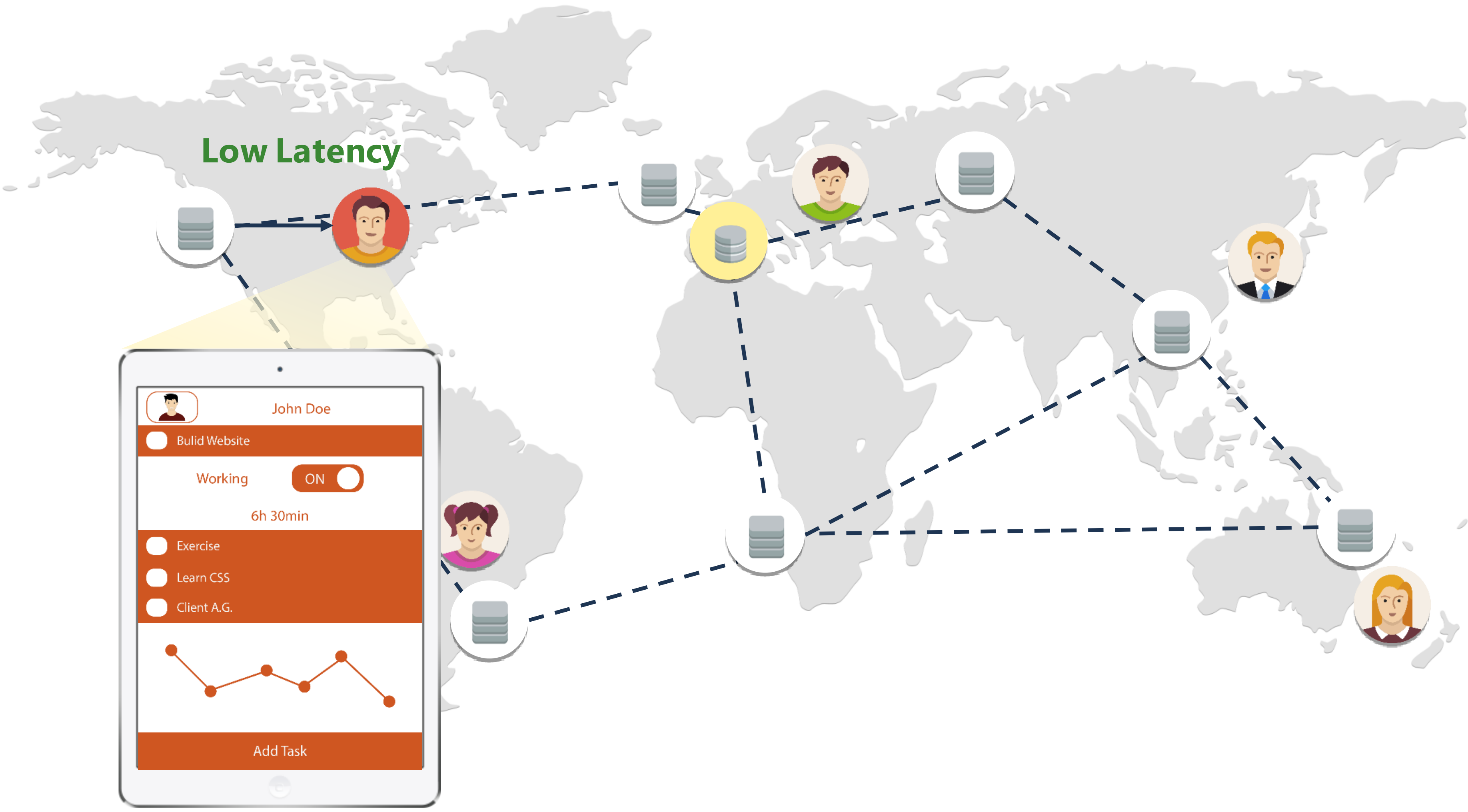
Low-Latency

Data is served by ubiquitous web-caches



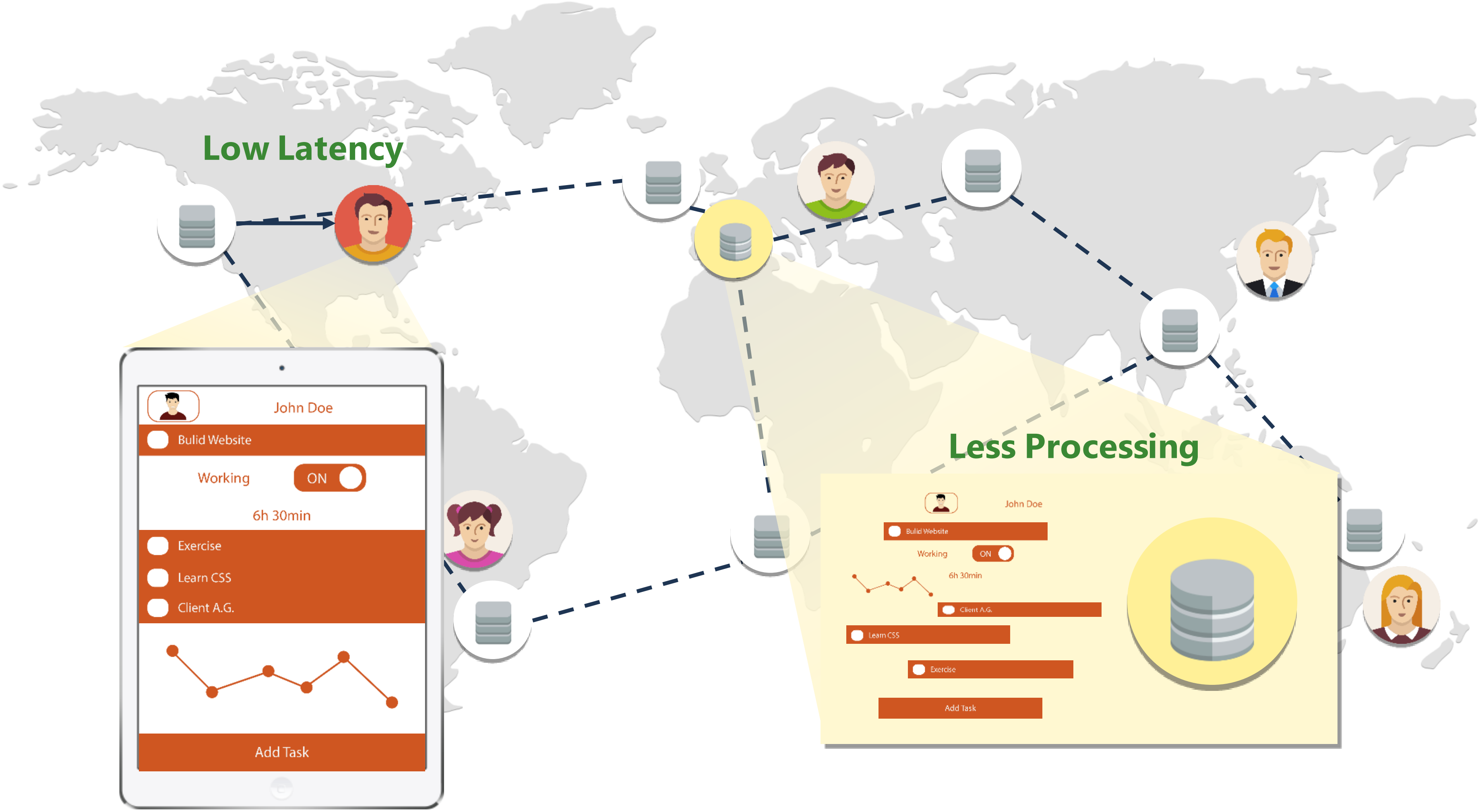
Low-Latency

Data is served by ubiquitous web-caches



Low-Latency

Data is served by ubiquitous web-caches



Low Latency

John Doe

Bulid Website

Working ON

6h 30min

Exercise

Learn CSS

Client A.G.

Add Task

Less Processing

John Doe

Bulid Website

Working ON

6h 30min

Client A.G.

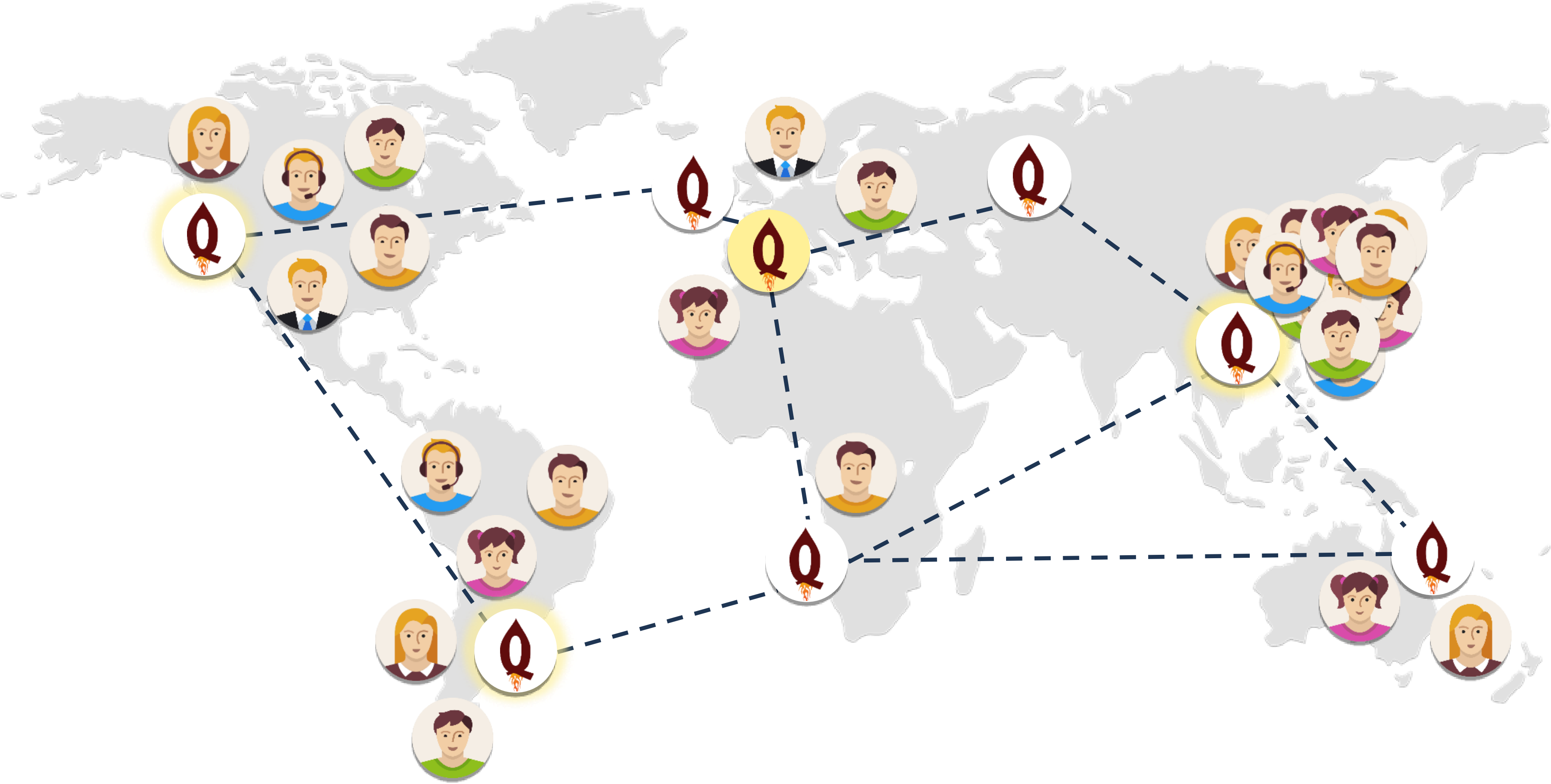
Learn CSS

Exercise

Add Task

Scaling

Scalable and highly available



Innovation

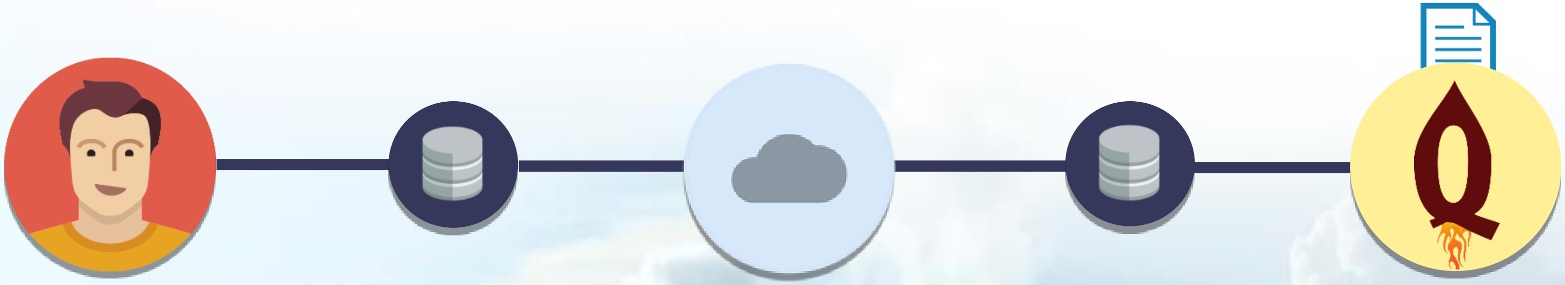
ASYNCHRONOUS
Decentralized
Approximate
Prefetching
BROADCAST
BUILDING
Memory
Scale
SPECTRAL
VIRTUALIZATION
PostgreSQL
DUPLICATES
INTELLIGENT
EVALUATION
SYSTEMS
PREDICTION
COMPARISON
SURVEY
CONTENT
Programming
INTERACTIVE
Alternative
VISUALIZATION
MODERN
CHARACTERISTICS
REPLICATION
BUSINESS
LIMITATIONS
Multitenant
USING
Simple
MODERATELY
NOSQL
OPTIMISTIC
OVERVIEW
DICTIONARIES
RELATIONALER
VALUE
DATASTORES
MODELING
DELIVERY
Performance
Service
look
CAPABILITIES
MEASUREMENT
TRANSACTIONS
FORBUILDING
LIGHTWEIGHT
MULTIMEDIA
OBJECT
Functions
DATASETS
ENABLING
causal
Live
MACHINE
WORKLOADS
challenges
DIMENSIONAL
ANALYSIS
OPERATION
PROVIDERS
Main
FILTERS
LATENCY
ISOLATION
IMPROVING
BENCHMARKING
History
DATA
RESOURCE
Column
HIGHLY
Approaches
AVAILABLE
Querying
PEER
Replication
DISTRIBUTED
HASHING
TOLERANT
Azure
CLASSIFICATION
NECESSITY
Storage
Taxonomy
LOCKING
STRONG
MECHANISMS
DATABASE
WEB
Internet
OPPORTUNITIES
LOAD
CLOUDBROWSER
PROCESSING
Concurrency
MASTERARBEIT
KEY
FREEING
PROVISIONING
computing
SYNCHRONIZATION
RESSOURCENVERWALTUNG
BLOOM
DATENBANKEN
Democratizing
APPLICATIONS
Software
LARGE
EVENTUALLY
STORES
Methods
techniques
UNDERSTANDING
STRUCTURES
CONSENSUS
Paxos
Replacement
STREAMING
Caching
MIGRATION
GENERATION
RELATIONAL
Authorization
MOBILE
PERSISTENCE
PERSPECTIVES
INSTRUMENTS
PARTITION
Consistency
ALGORITHMS
ANALYTICS
ANFORDERUNGEN
IMPLEMENTATIONS
PROXY
Enterprise
Engineering
Changed
NETWORK
High
Comparative
Snapshot
MULTIPOINT
SQL
POLICY
PRACTICAL
Solutions
SERVERS
CLUSTERING
BASED
ENGAGEMENTS
AWARE
Dynamics
CAP
CASE
SERIALIZABLE
SEARCHING
PROBABILISTICALLY
ANWENDUNGEN
PRESENTATIN
MAPREDUCE
INVESTIGATION
mapping
CLOUD
Cassandra
Multi
delivering
MANAGEMENT
CONTROL
Platforms
TENANT
OVERCOMING
Maintenance
MAKING
Shared
ARCHITECTURES
BLOOMFILTER
PUBLICATION
ENVIRONMENTS
SCALABLE
SUPPORTING
TECHNOLOGY
TREE
BALANCING
SCHEMA
LEARNING
ORIENTED
Datenbanksysteme
ELASTIC
GOOGLES
TRANSAKTIONSUNTERSTÜTZUNG
HETEROGENEOUS
Theory
FRAMEWORK
FREQUENCY
Feasibility

Innovation

Problem: changes cause stale data

5 Years
Research & Development

α **New Algorithms**
Solve Consistency Problem

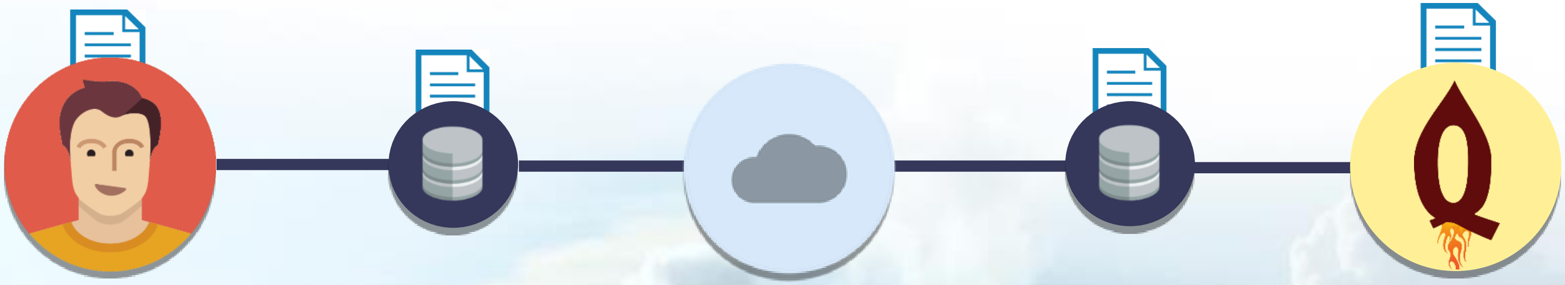


Innovation

Problem: changes cause stale data

5 Years
Research & Development

New Algorithms
Solve Consistency Problem



Innovation

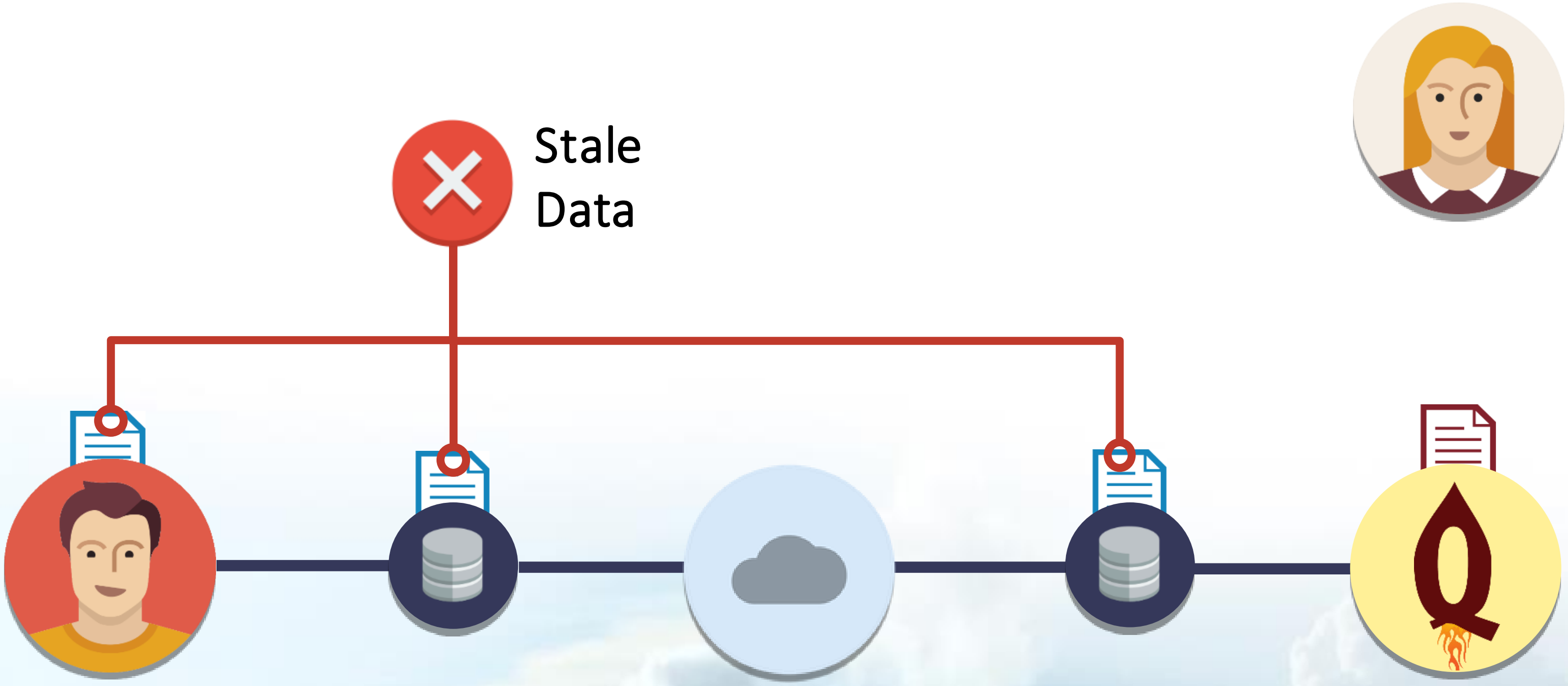
Problem: changes cause stale data

5 Years

Research & Development

New Algorithms

Solve Consistency Problem



Innovation

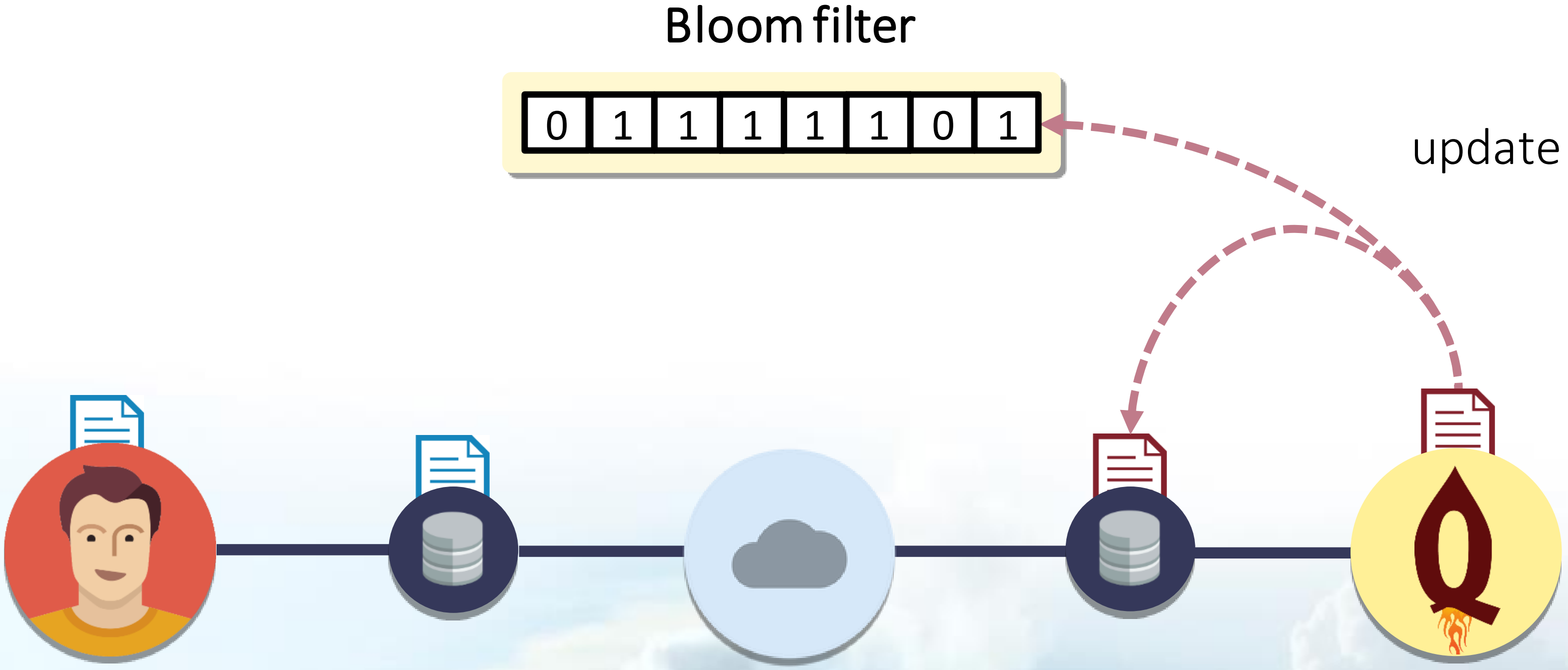
Solution: Baqend proactively revalidates data

5 Years

Research & Development

New Algorithms

Solve Consistency Problem

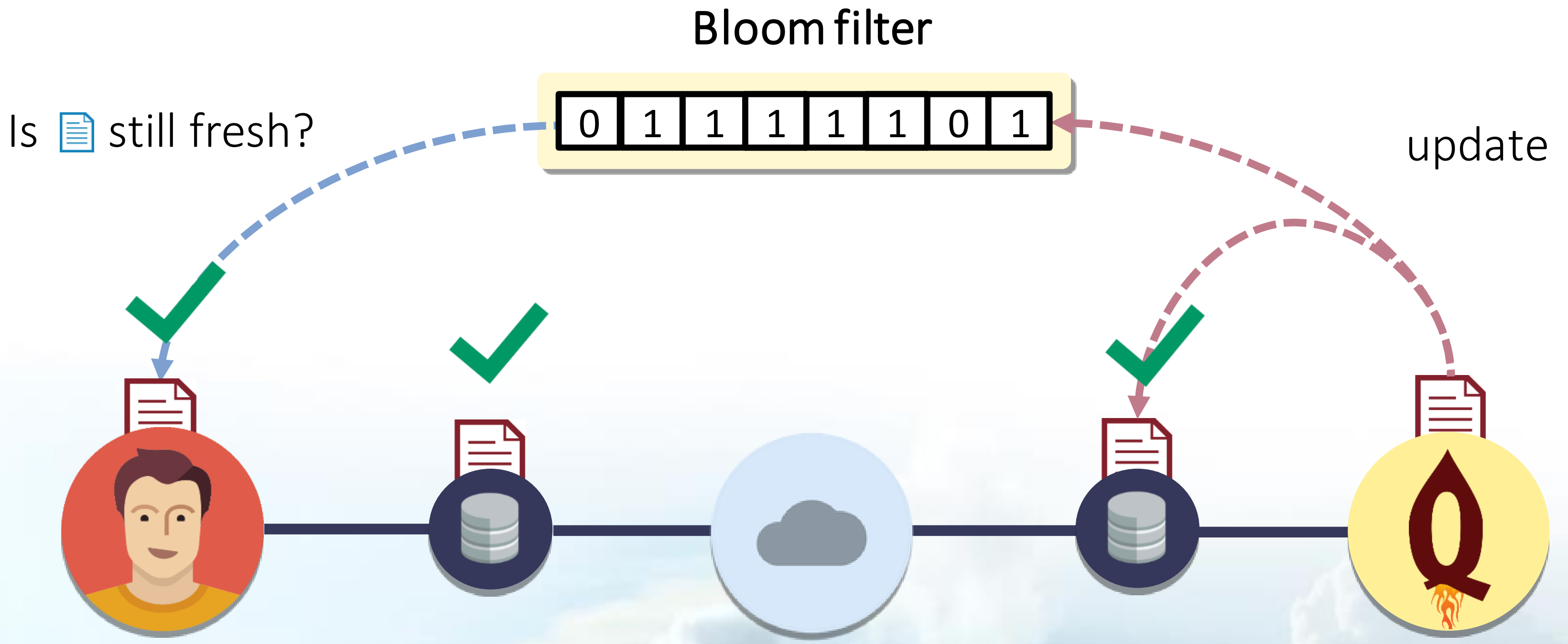


Innovation

Solution: Baqend proactively revalidates data

5 Years
Research & Development

New Algorithms
Solve Consistency Problem



Bloom filter

0	1	1	1	1	1	0	1

Is 📄 still fresh?

update

Innovation

Solution: Baqend proactively revalidates data

F. Gessert, F. Bücklers, und N. Ritter, „ORESTES: a Scalable Database-as-a-Service Architecture for Low Latency“, in *CloudDB 2014*, 2014.

F. Gessert und F. Bücklers, „ORESTES: ein System für horizontal skalierbaren Zugriff auf Cloud-Datenbanken“, in Informatiktage 2013, 2013.

F. Gessert und F. Bücklers, *Performanz- und Reaktivitätssteigerung von OODBMS vermittels der Web-Caching-Hierarchie*. Bachelorarbeit, 2010.

M. Schaarschmidt, F. Gessert, und N. Ritter, „Towards Automated Polyglot Persistence“, in BTW 2015.

S. Friedrich, W. Wingerath, F. Gessert, und N. Ritter, „NoSQL OLTP Benchmarking: A Survey“, in *44. Jahrestagung der Gesellschaft für Informatik*, 2014, Bd. 232, S. 693–704.

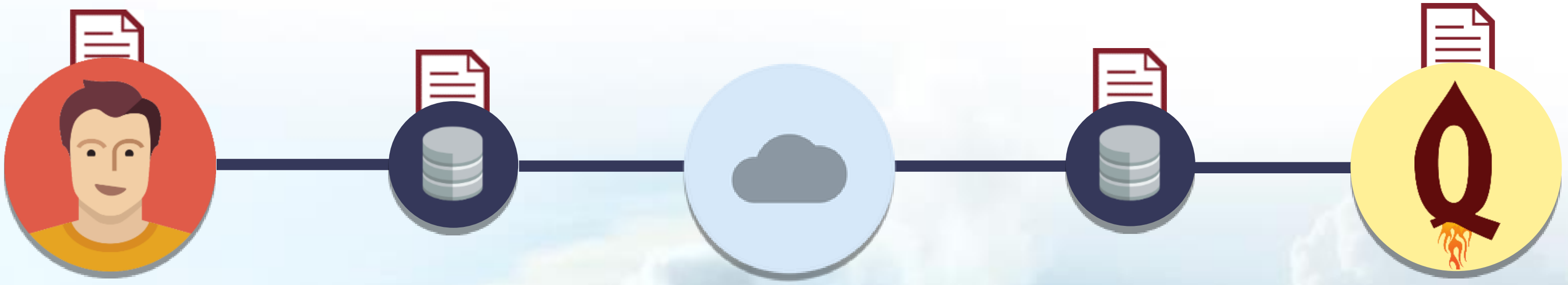
F. Gessert, S. Friedrich, W. Wingerath, M. Schaarschmidt, und N. Ritter, „Towards a Scalable and Unified REST API for Cloud Data Stores“, in *44. Jahrestagung der GI*, Bd. 232, S. 723–734.

F. Gessert, M. Schaarschmidt, W. Wingerath, S. Friedrich, und N. Ritter, „The Cache Sketch: Revisiting Expiration-based Caching in the Age of Cloud Data Management“, in BTW 2015.

F. Gessert und F. Bücklers, *Kohärentes Web-Caching von Datenbankobjekten im Cloud Computing*. Masterarbeit 2012.

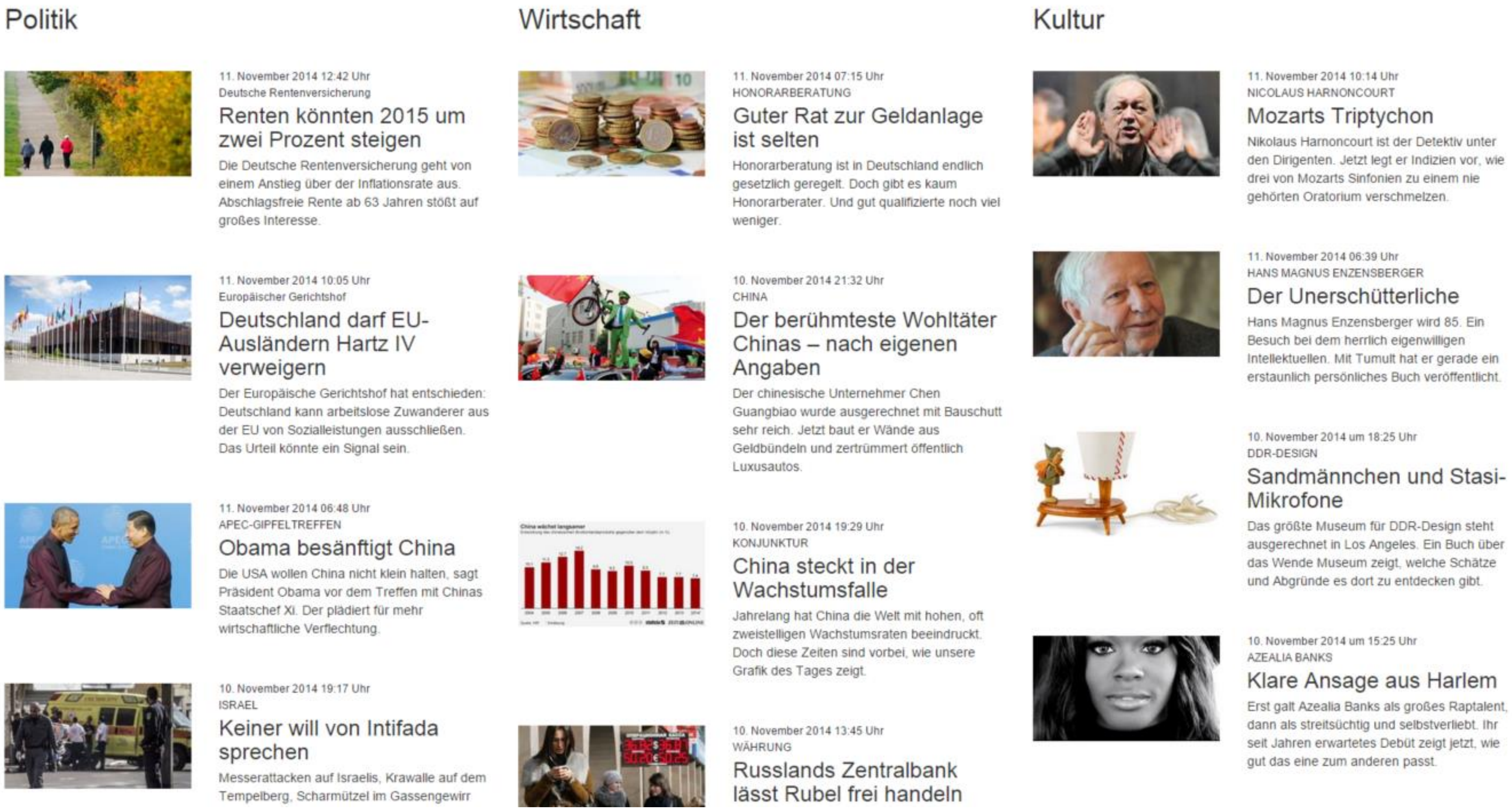
W. Wingerath, S. Friedrich, und F. Gessert, „Who Watches the Watchmen? On the Lack of Validation in NoSQL Benchmarking“, in BTW 2015.

F. Gessert, „Skalierbare NoSQL- und Cloud-Datenbanken in Forschung und Praxis“, BTW 2015



Page-Load Times

What impact does caching have in practice?



Politik

11. November 2014 12:42 Uhr
Deutsche Rentenversicherung

Renten könnten 2015 um zwei Prozent steigen

Die Deutsche Rentenversicherung geht von einem Anstieg über der Inflationsrate aus. Abschlagsfreie Rente ab 63 Jahren stößt auf großes Interesse.

11. November 2014 10:05 Uhr
Europäischer Gerichtshof

Deutschland darf EU-Ausländern Hartz IV verweigern

Der Europäische Gerichtshof hat entschieden: Deutschland kann arbeitslose Zuwanderer aus der EU von Sozialleistungen ausschließen. Das Urteil könnte ein Signal sein.

11. November 2014 06:48 Uhr
APEC-GIPFELTREFFEN

Obama besänftigt China

Die USA wollen China nicht klein halten, sagt Präsident Obama vor dem Treffen mit Chinas Staatschef Xi. Der plädiert für mehr wirtschaftliche Verflechtung.

10. November 2014 19:17 Uhr
ISRAEL

Keiner will von Intifada sprechen

Messerattacken auf Israelis, Krawalle auf dem Tempelberg, Scharmützel im Gassengewirr

Wirtschaft

11. November 2014 07:15 Uhr
HONORARBERATUNG

Guter Rat zur Geldanlage ist selten

Honorarberatung ist in Deutschland endlich gesetzlich geregelt. Doch gibt es kaum Honorarberater. Und gut qualifizierte noch viel weniger.

10. November 2014 21:32 Uhr
CHINA

Der berühmteste Wohltäter Chinas – nach eigenen Angaben

Der chinesische Unternehmer Chen Guangbiao wurde ausgerechnet mit Bauschutt sehr reich. Jetzt baut er Wände aus Geldbündeln und zertrümmert öffentlich Luxusautos.

10. November 2014 19:29 Uhr
KONJUNKTUR

China steckt in der Wachstumsfalle

Jahrelang hat China die Welt mit hohen, oft zweistelligen Wachstumsraten beeindruckt. Doch diese Zeiten sind vorbei, wie unsere Grafik des Tages zeigt.

10. November 2014 13:45 Uhr
WÄHRUNG

Russlands Zentralbank lässt Rubel frei handeln

Kultur

11. November 2014 10:14 Uhr
NICOLAUS HARNONCOURT

Mozarts Triptychon

Nikolaus Harnoncourt ist der Detektiv unter den Dirigenten. Jetzt legt er Indizien vor, wie drei von Mozarts Sinfonien zu einem nie gehörten Oratorium verschmelzen.

11. November 2014 06:39 Uhr
HANS MAGNUS ENZENSBERGER

Der Unerschütterliche

Hans Magnus Enzensberger wird 85. Ein Besuch bei dem herrlich eigenwilligen Intellektuellen. Mit Tumult hat er gerade ein erstaunlich persönliches Buch veröffentlicht.

10. November 2014 um 18:25 Uhr
DDR-DESIGN

Sandmännchen und Stasi-Mikrofone

Das größte Museum für DDR-Design steht ausgerechnet in Los Angeles. Ein Buch über das Wende Museum zeigt, welche Schätze und Abgründe es dort zu entdecken gibt.

10. November 2014 um 15:25 Uhr
AZEALIA BANKS

Klare Ansage aus Harlem

Erst galt Azealia Banks als großes Raptalent, dann als streitsüchtig und selbstverliebt. Ihr seit Jahren erwartetes Debüt zeigt jetzt, wie gut das eine zum anderen passt.

Page-Load Times

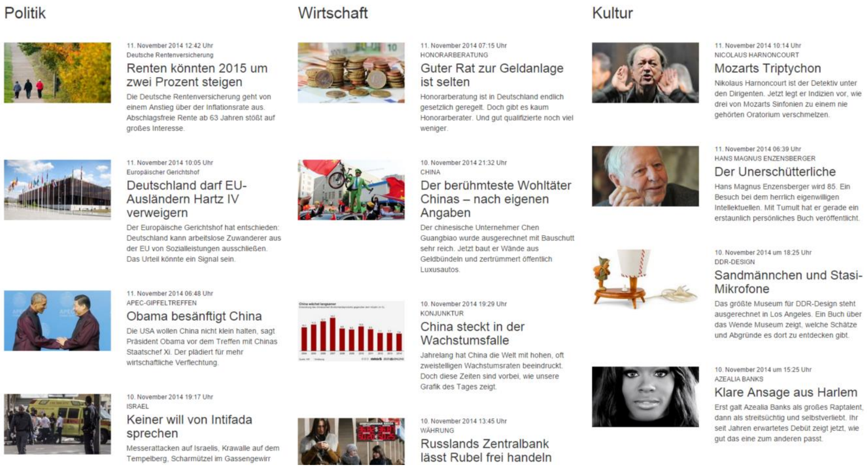
What impact does caching have in practice?

Parse

kinvey

Windows Azure

APiOMAT



Politik

11. November 2014 12:42 Uhr
Deutsche Rentenversicherung
Renten könnten 2015 um zwei Prozent steigen
Die Deutsche Rentenversicherung geht von einem Anstieg über der Inflationsrate aus. Abschlagsfreie Rente ab 63 Jahren stößt auf großes Interesse.

11. November 2014 10:05 Uhr
Europäischer Gerichtshof
Deutschland darf EU-Ausländern Hartz IV verweigern
Der Europäische Gerichtshof hat entschieden: Deutschland kann arbeitslose Zuwanderer aus der EU von Sozialleistungen ausschließen. Das Urteil könnte ein Signal sein.

11. November 2014 06:48 Uhr
APEC-GIPFELTREFFEN
Obama besänftigt China
Die USA wollen China nicht klein halten, sagt Präsident Obama vor dem Treffen mit Chinas Staatschef Xi. Der plädiert für mehr wirtschaftliche Verflechtung.

10. November 2014 19:17 Uhr
ISRAEL
Keiner will von Intifada sprechen
Messerattacken auf Israelis, Krawalle auf dem Tempelberg, Scharmützel im Gassengewirr

Wirtschaft

11. November 2014 07:15 Uhr
HONORARBERATUNG
Guter Rat zur Geldanlage ist selten
Honorarberatung ist in Deutschland endlich gesetzlich geregelt. Doch gibt es kaum Honorarberater. Und gut qualifizierte noch viel weniger.

10. November 2014 21:32 Uhr
CHINA
Der berühmteste Wohltäter Chinas – nach eigenen Angaben
Der chinesische Unternehmer Chen Guangbiao wurde ausgerechnet mit Bauschutt sehr reich. Jetzt baut er Wände aus Geldbündeln und zertrümmert öffentlich Luxusautos.

10. November 2014 19:29 Uhr
KONJUNKTUR
China steckt in der Wachstumsfalle
Jahrelang hat China die Welt mit hohen, oft zweistelligen Wachstumsraten beeindruckt. Doch diese Zeiten sind vorbei, wie unsere Grafik des Tages zeigt.

10. November 2014 13:45 Uhr
WÄHRUNG
Russlands Zentralbank lässt Rubel frei handeln

Kultur

11. November 2014 10:14 Uhr
NICOLAUS HARNONCOURT
Mozarts Triptychon
Nikolaus Harnoncourt ist der Detektiv unter den Dirigenten. Jetzt legt er Indizien vor, wie drei von Mozarts Sinfonien zu einem nie gehörten Oratorium verschmelzen.

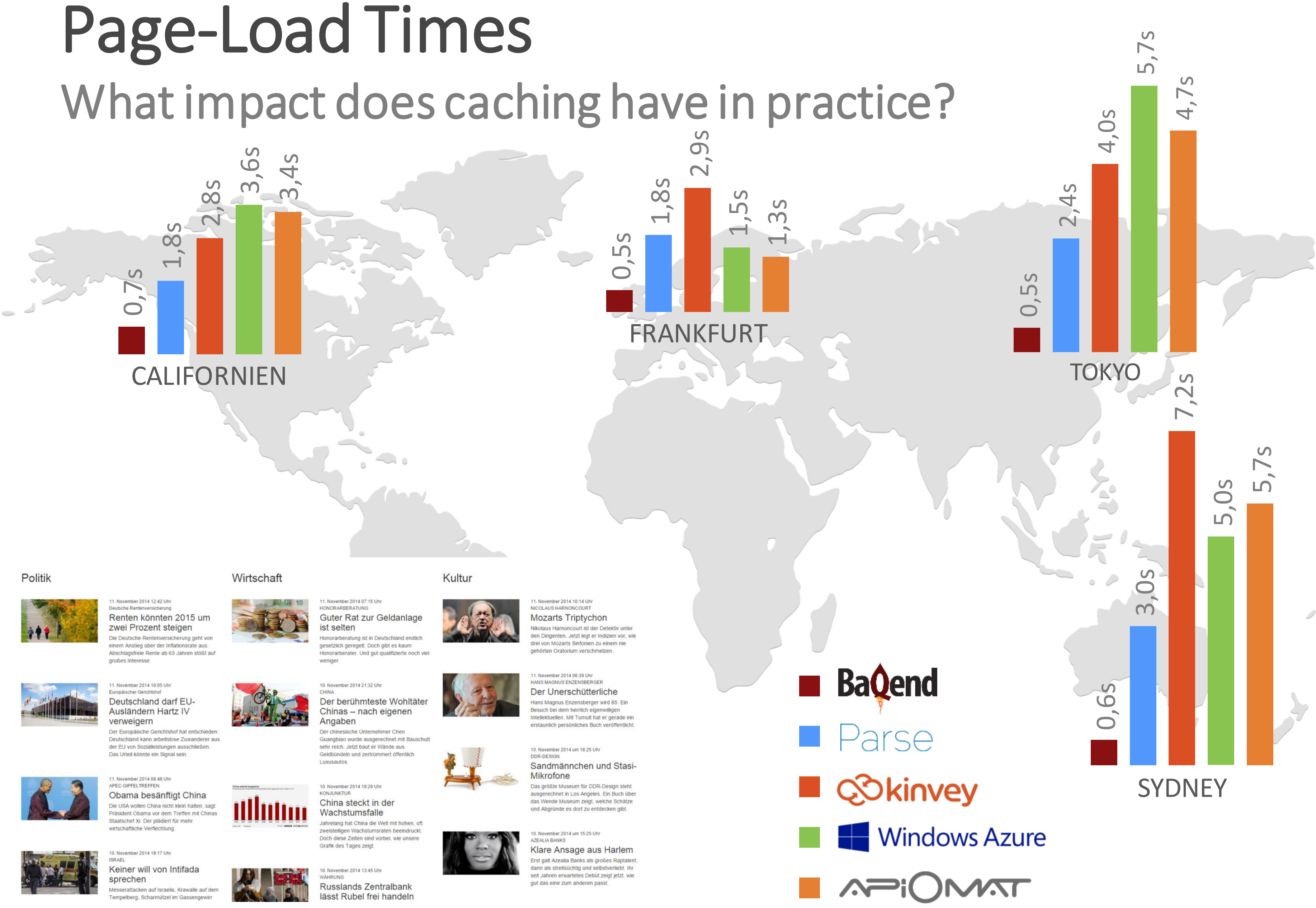
11. November 2014 06:39 Uhr
HANS MAGNUS ENZENSBERGER
Der Unerschütterliche
Hans Magnus Enzensberger wird 85. Ein Besuch bei dem herrlich eigenwilligen Intellektuellen. Mit Tumult hat er gerade ein erstaunlich persönliches Buch veröffentlicht.

10. November 2014 um 18:25 Uhr
DDR-DESIGN
Sandmännchen und Stasi-Mikrofone
Das größte Museum für DDR-Design steht ausgerechnet in Los Angeles. Ein Buch über das Wende Museum zeigt, welche Schätze und Abgründe es dort zu entdecken gibt.

10. November 2014 um 15:25 Uhr
AZEALIA BANKS
Klare Ansage aus Harlem
Erst galt Azealia Banks als großes Raptalent, dann als streitsüchtig und selbstverliebt. Ihr seit Jahren erwartetes Debüt zeigt jetzt, wie gut das eine zum anderen passt.

Page-Load Times

What impact does caching have in practice?



Politik

11. November 2014 12:42 Uhr
Deutsche Rentenversicherung
Renten könnten 2015 um zwei Prozent steigen
Die Deutsche Rentenversicherung geht von einem Anstieg über der Inflationsrate aus. Abschlagsfreie Rente ab 63 Jahren stößt auf großes Interesse.

11. November 2014 10:05 Uhr
Europäischer Gerichtshof
Deutschland darf EU-Ausländern Hartz IV verweigern
Der Europäische Gerichtshof hat entschieden: Deutschland kann arbeitslose Zuwanderer aus der EU von Sozialleistungen ausschließen. Das Urteil könnte ein Signal sein.

11. November 2014 06:48 Uhr
APEC-GIPFELTREFFEN
Obama besänftigt China
Die USA wollen China nicht klein halten, sagt Präsident Obama vor dem Treffen mit Chinas Staatschef Xi. Der plädiert für mehr wirtschaftliche Verflechtung.

10. November 2014 19:17 Uhr
ISRAEL
Keiner will von Intifada sprechen
Messerattacken auf Israelis, Krawalle auf dem Tempelberg, Scharmützel im Gassengewirr

Wirtschaft

11. November 2014 07:15 Uhr
HONORARBERATUNG
Guter Rat zur Geldanlage ist selten
Honorarberatung ist in Deutschland endlich gesetzlich geregelt. Doch gibt es kaum Honorarberater. Und gut qualifizierte noch viel weniger.

10. November 2014 21:32 Uhr
CHINA
Der berühmteste Wohltäter Chinas – nach eigenen Angaben
Der chinesische Unternehmer Chen Guangbiao wurde ausgerechnet mit Bauschutt sehr reich. Jetzt baut er Wände aus Geldbündeln und zertrümmert öffentlich Luxusautos.

10. November 2014 19:29 Uhr
KONJUNKTUR
China steckt in der Wachstumsfalle
Jahrelang hat China die Welt mit hohen, oft zweistelligen Wachstumsraten beeindruckt. Doch diese Zeiten sind vorbei, wie unsere Grafik des Tages zeigt.

10. November 2014 13:45 Uhr
WÄHRUNG
Russlands Zentralbank lässt Rubel frei handeln

Kultur

11. November 2014 10:14 Uhr
NICOLAUS HARNONCOURT
Mozarts Triptychon
Nikolaus Harnoncourt ist der Detektiv unter den Dirigenten. Jetzt legt er Indizien vor, wie drei von Mozarts Sinfonien zu einem nie gehörten Oratorium verschmelzen.

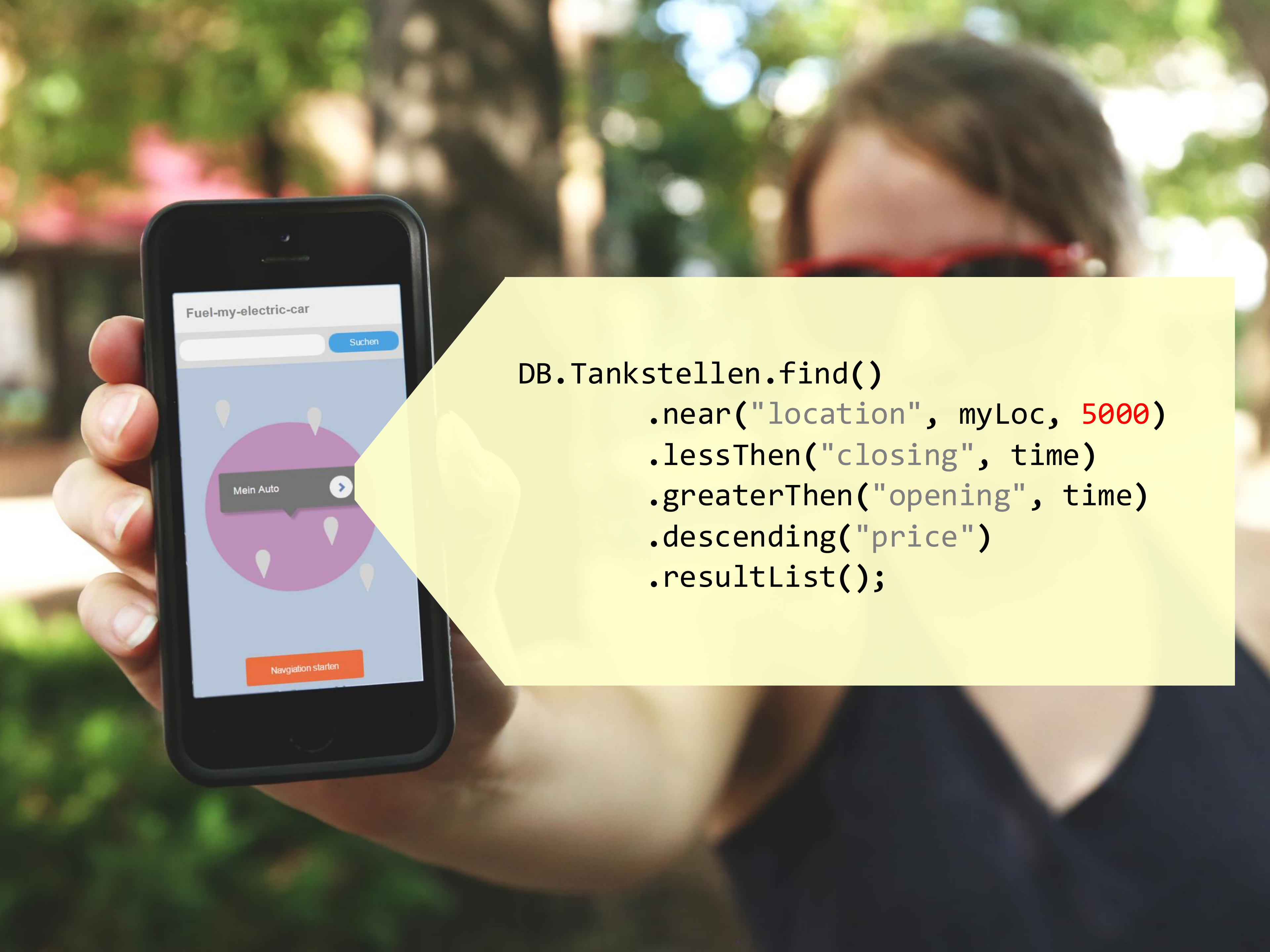
11. November 2014 06:39 Uhr
HANS MAGNUS ENZENSBERGER
Der Unerschütterliche
Hans Magnus Enzensberger wird 85. Ein Besuch bei dem herrlich eigenwilligen Intellektuellen. Mit Tumult hat er gerade ein erstaunlich persönliches Buch veröffentlicht.

10. November 2014 um 18:25 Uhr
DDR-DESIGN
Sandmännchen und Stasi-Mikrofone
Das größte Museum für DDR-Design steht ausgerechnet in Los Angeles. Ein Buch über das Wende Museum zeigt, welche Schätze und Abgründe es dort zu entdecken gibt.

10. November 2014 um 15:25 Uhr
AZEALIA BANKS
Klare Ansage aus Harlem
Erst galt Azealia Banks als großes Raptalent, dann als streitsüchtig und selbstverliebt. Ihr seit Jahren erwartetes Debüt zeigt jetzt, wie gut das eine zum anderen passt.

How is this used from a **develeoper's** perspective?

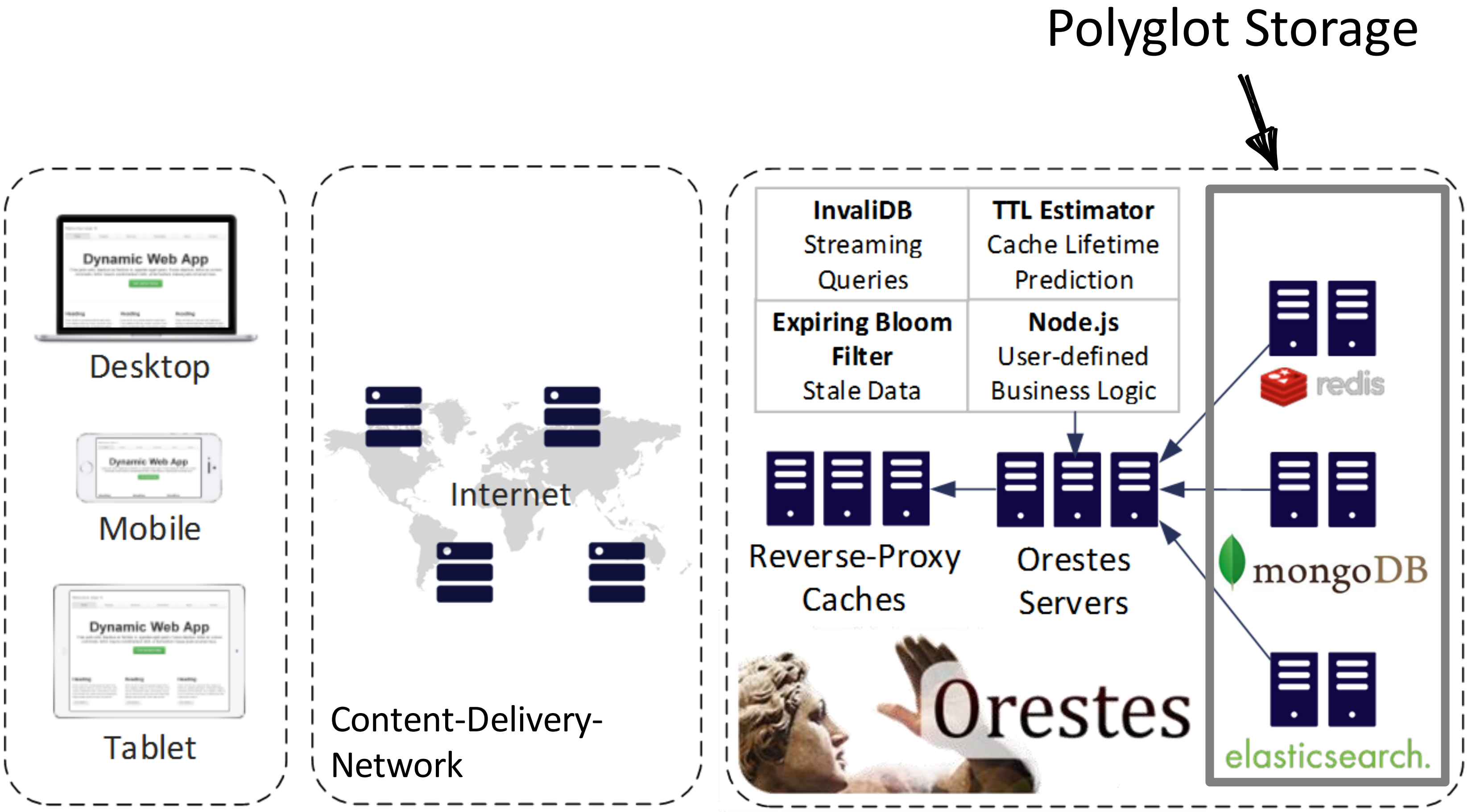
Backend-as-a-**Service**



```
DB.Tankstellen.find()
         .near("location", myLoc, 5000)
         .lessThen("closing", time)
         .greaterThen("opening", time)
         .descending("price")
         .resultList();
```

Baqend Architecture

Our Infrastructure



Polyglot Storage

Desktop

Mobile

Tablet

Internet

Content-Delivery-Network

InvaliDB
Streaming Queries

TTL Estimator
Cache Lifetime Prediction

Expiring Bloom Filter
Stale Data

Node.js
User-defined Business Logic

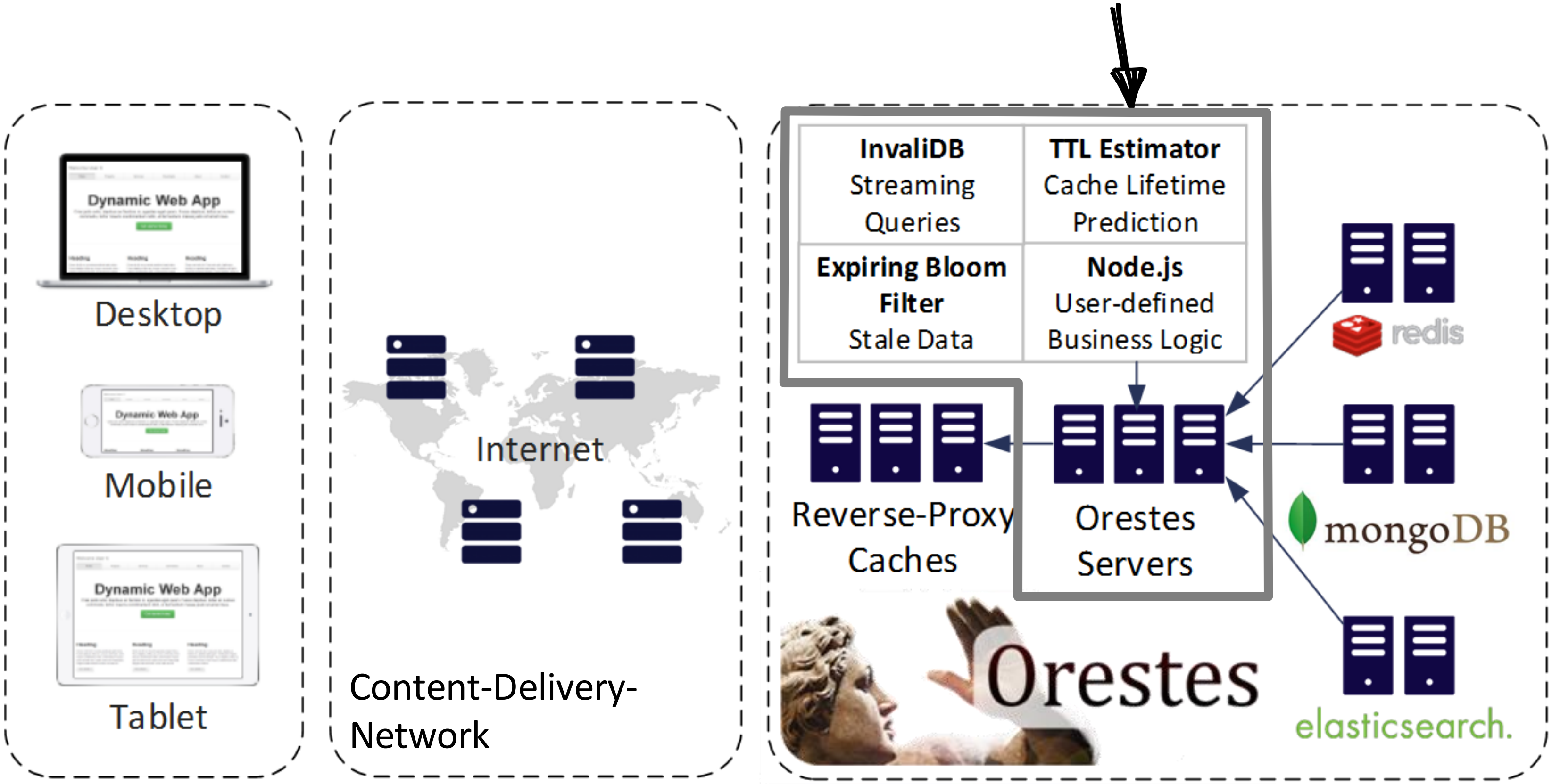
Reverse-Proxy Caches

Orestes Servers

Baqend Architecture

Our Infrastructure

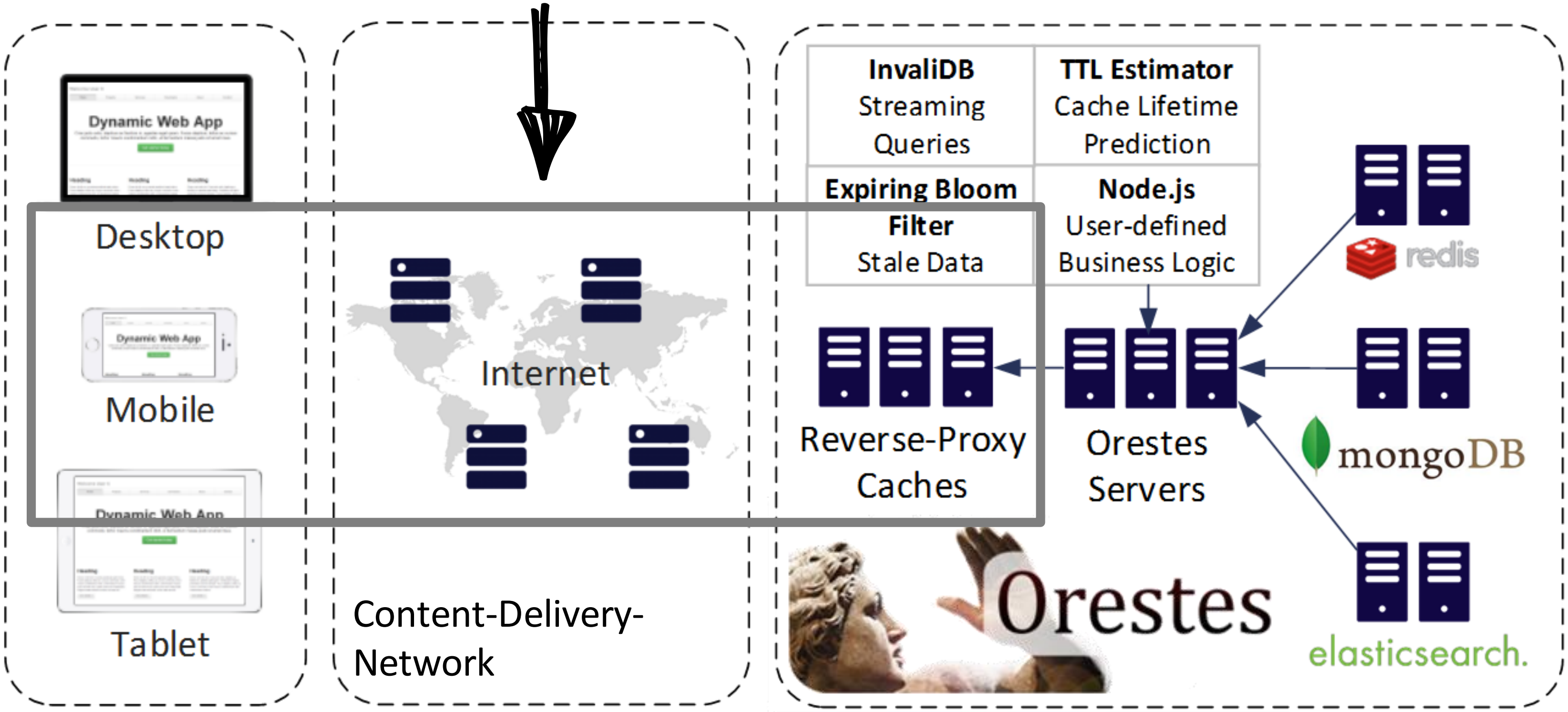
Database-as-a-Service Middleware:
Caching, Transactions, Schemas,
Invalidation Detection, …



Baqend Architecture

Our Infrastructure

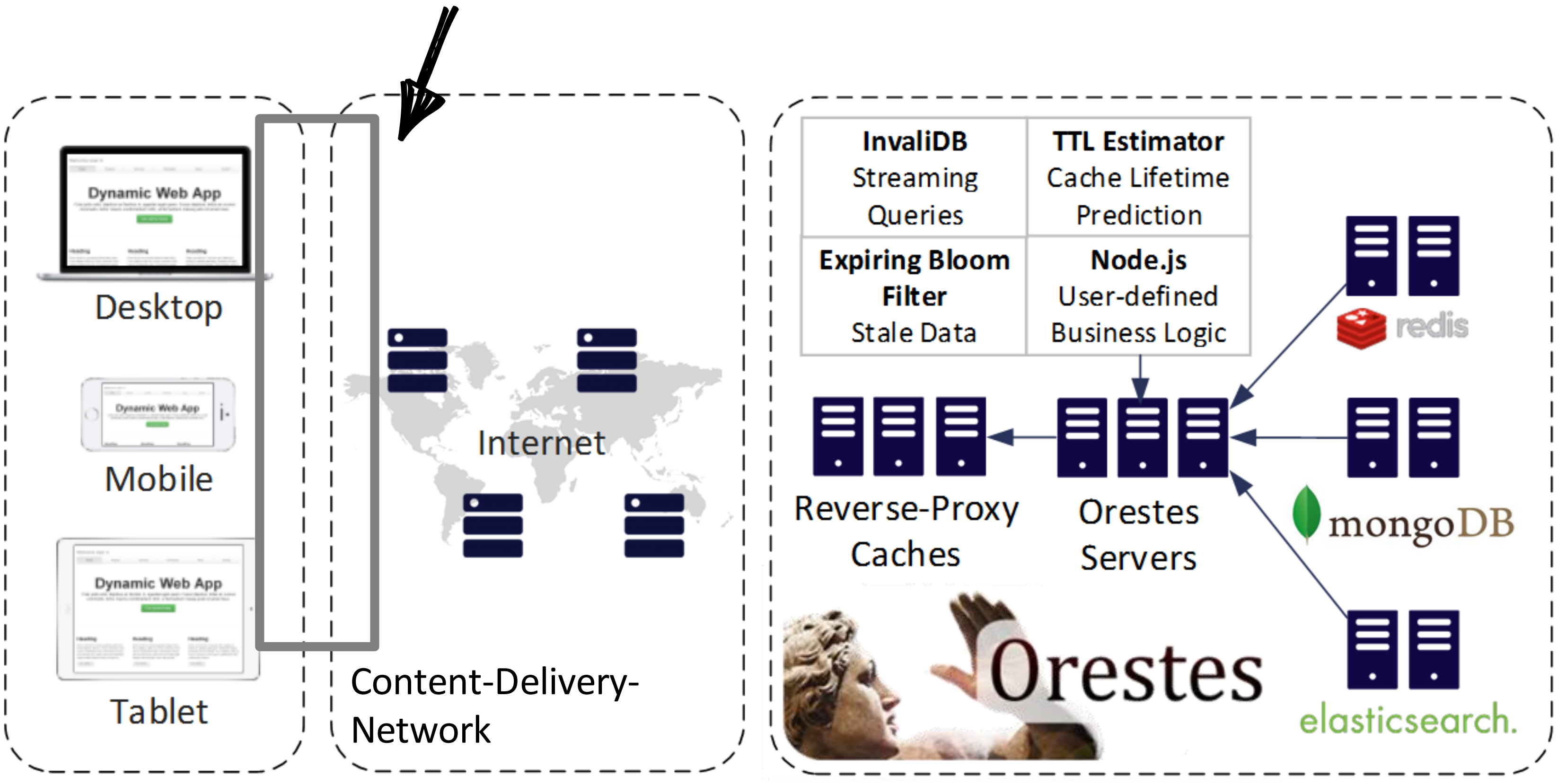
Standard HTTP Caching



Baqend Architecture

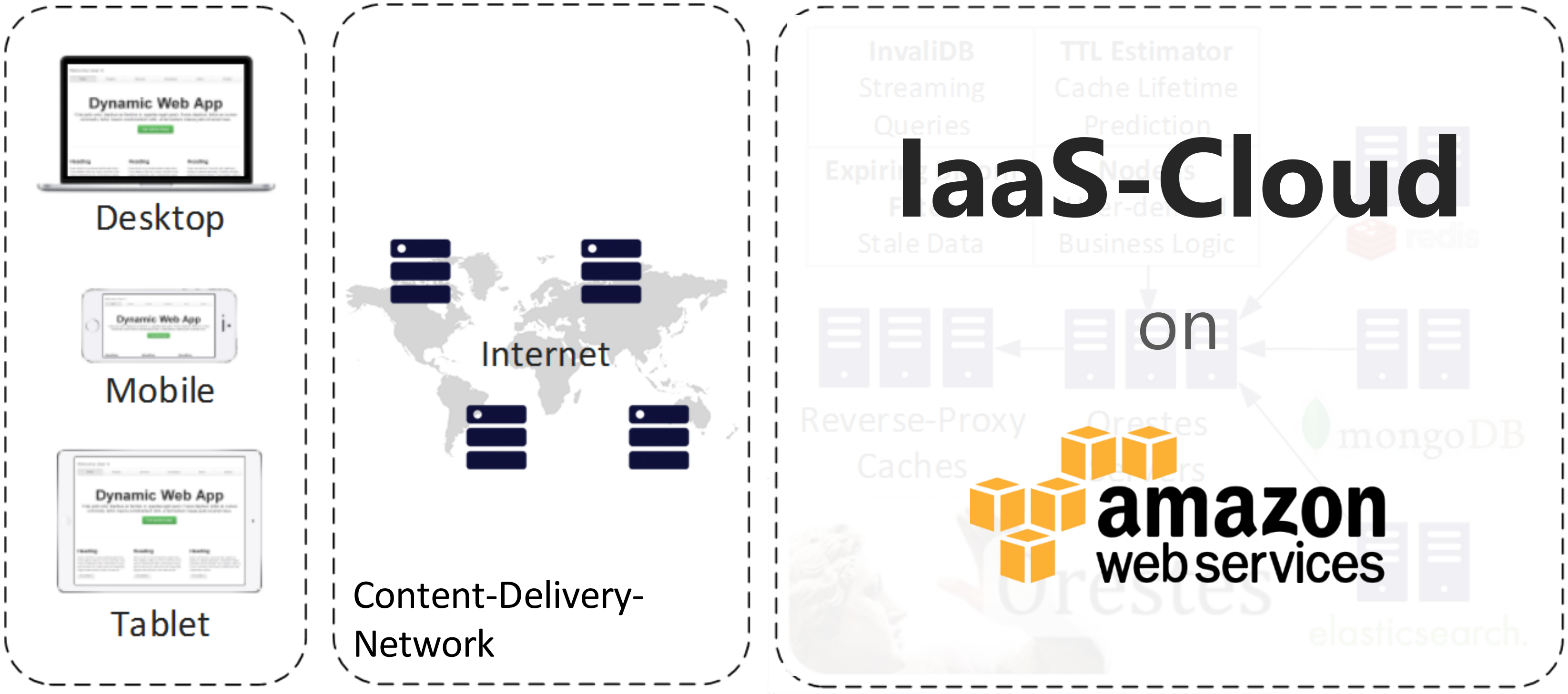
Our Infrastructure

Unified REST API



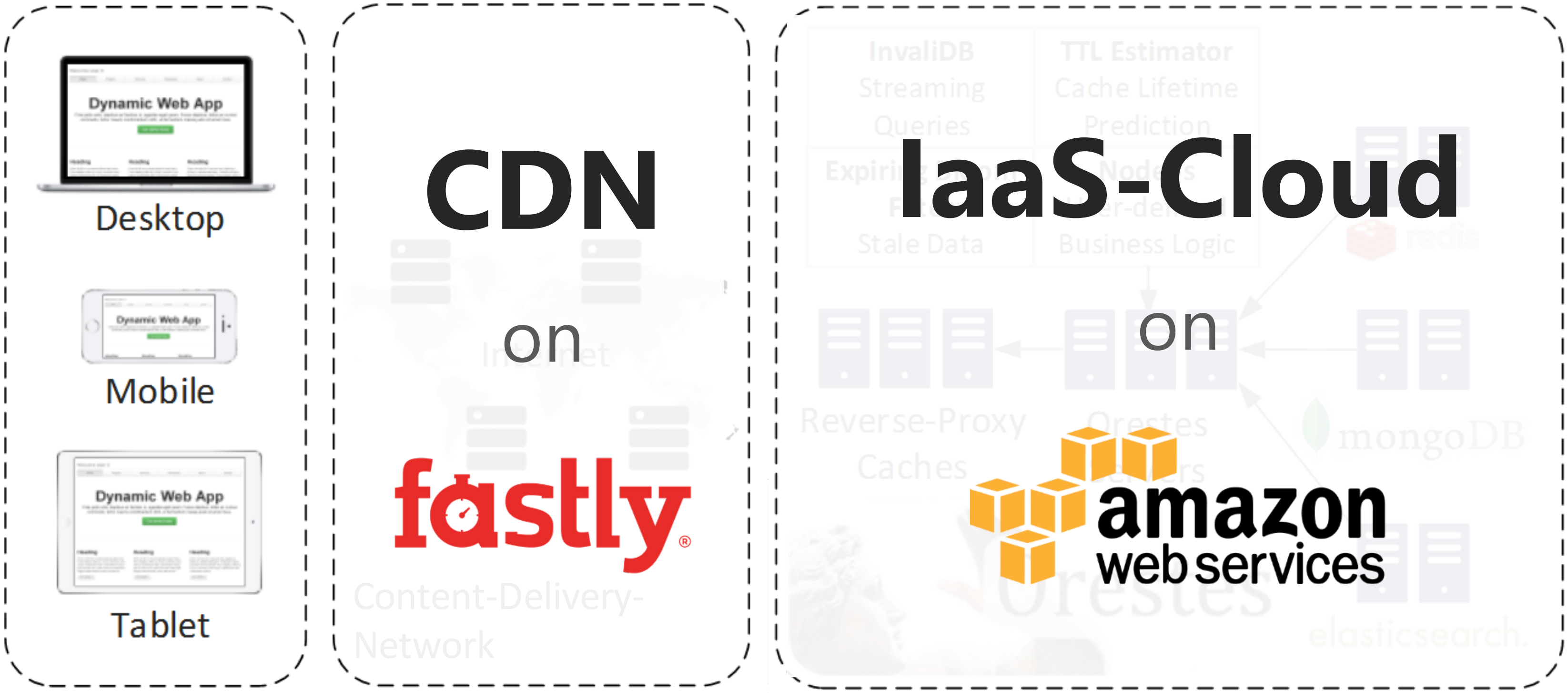
Baqend Architecture

Our Infrastructure



Baqend Architecture

Our Infrastructure



Desktop

Mobile

Tablet

CDN

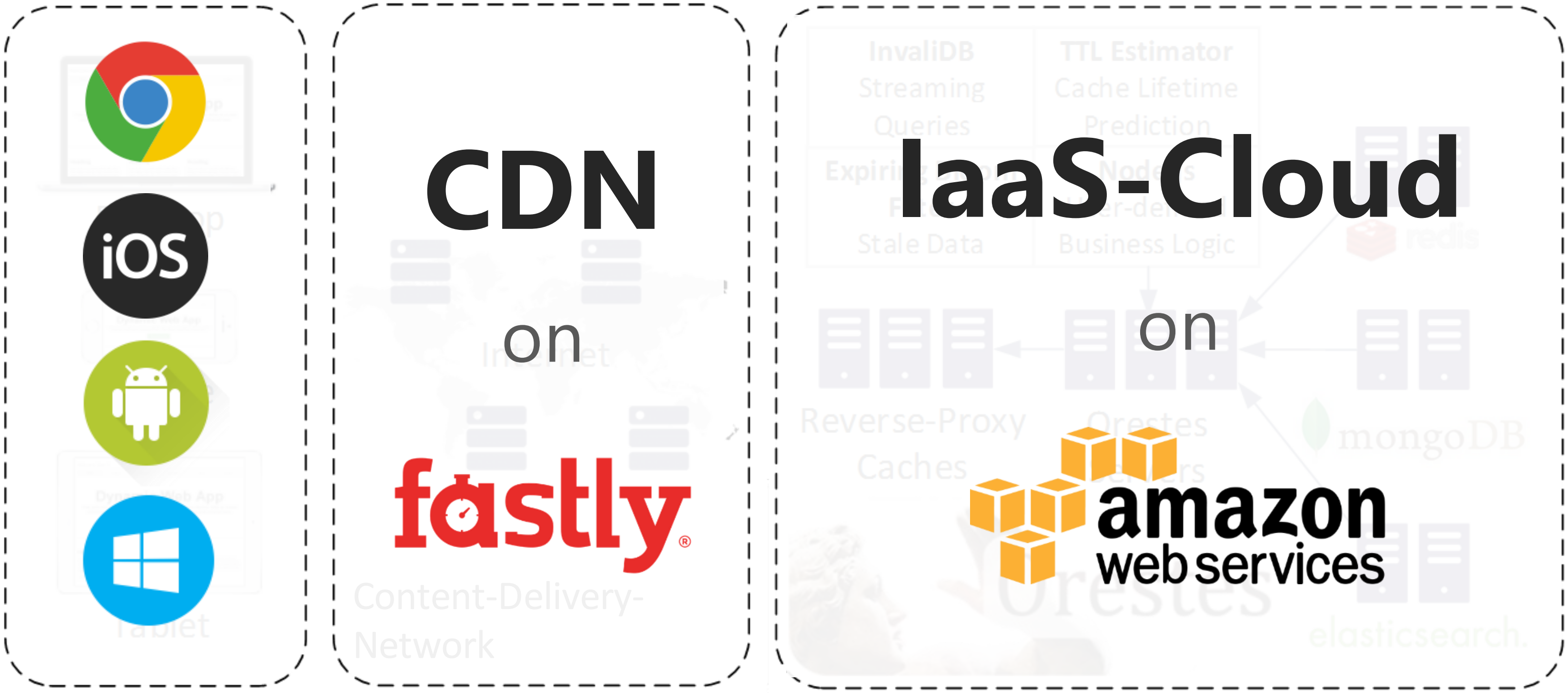
on

IaaS-Cloud

on

Baqend Architecture

Our Infrastructure

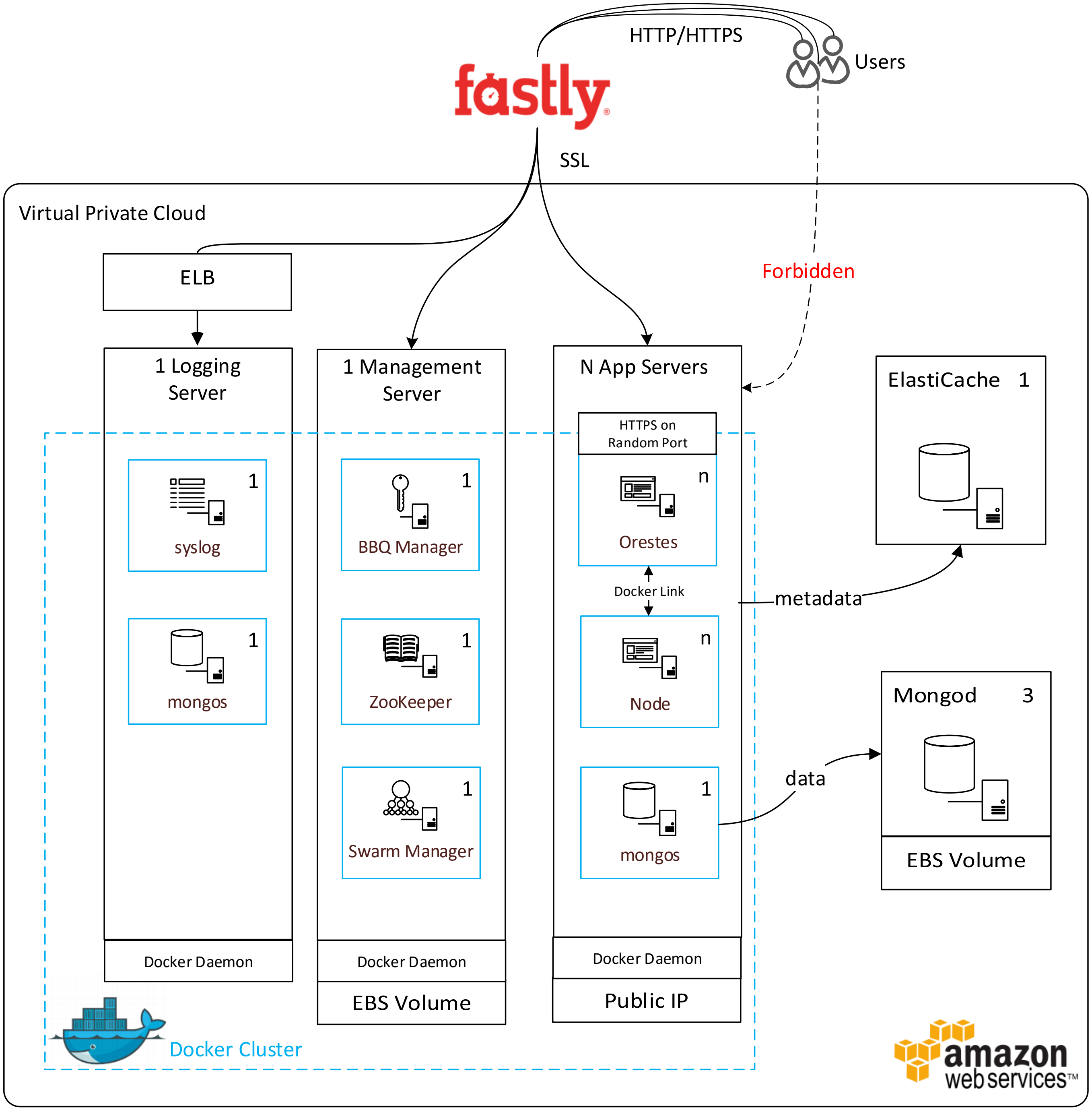


CDN

on

IaaS-Cloud

on


HTTP/HTTPS
Users
fastly
SSL
Virtual Private Cloud
Forbidden
ELB
1 Logging Server
1 Management Server
N App Servers
ElastiCache 1
HTTPS on Random Port
syslog 1
BBQ Manager 1
Orestes n
Docker Link
metadata
mongos 1
ZooKeeper 1
Node n
Mongod 3
data
Swarm Manager 1
mongos 1
EBS Volume
Docker Daemon
Docker Daemon
Docker Daemon
EBS Volume
Public IP
Docker Cluster
amazon web services™

AWS Services

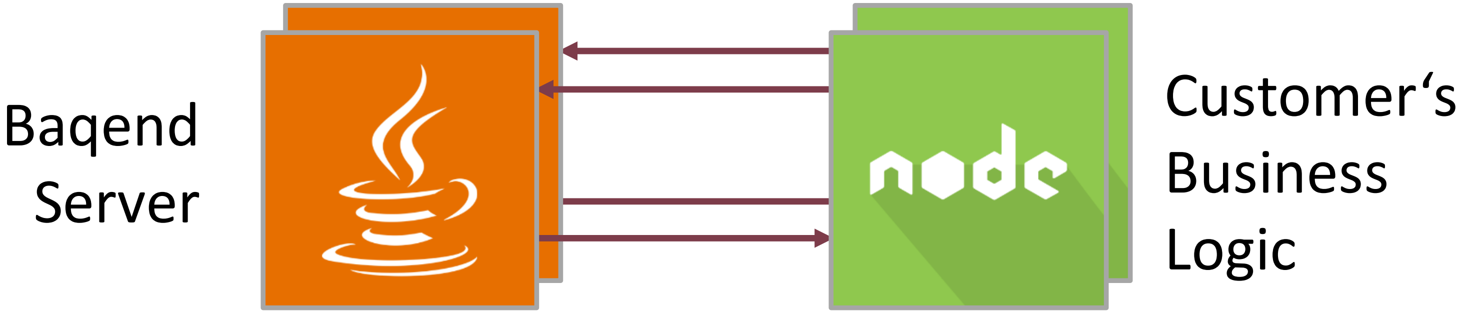
Services we use

- Route 53, EC2, ASGs, IAM etc.
- **Elastic Load Balancer:** TCP Balancing for Logging
  - Not suited for multi-tenant SSL termination: ELB cannot dynamically route to an IP:port pair
- **Redis ElastiCache**: Metadata Storage
  - Easy to use but very limited: no Redis cluster support, no append-only files, bad snapshotting
- **What we don‘t use:**
  - **Beanstalk:** supports Docker but needs a dedicated EC2 instance
  - **Cloudfront:** useless invalidations, expensive
  - **DynamoDB**: difficult to scale, very limited queries

Containerization

Why we need containers & cluster management

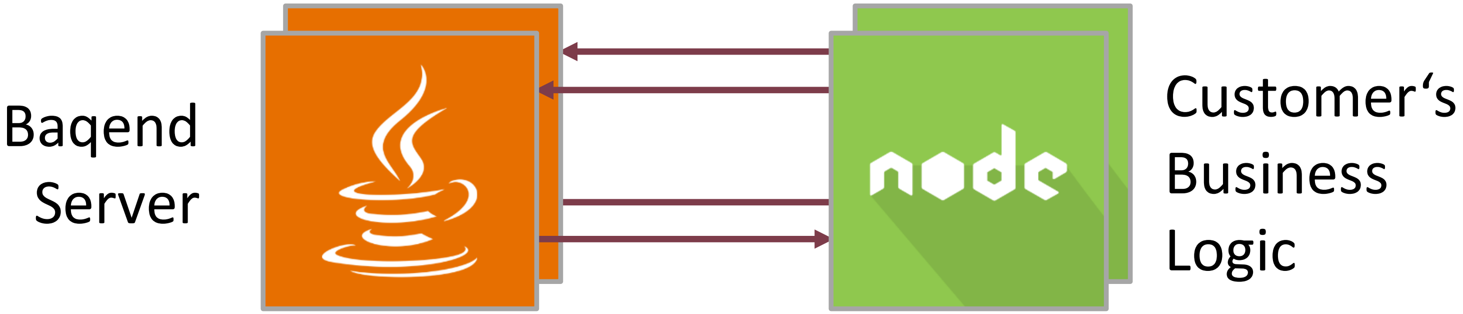
- Every tenant needs a private JVM and Node.JS process



Containerization

Why we need containers & cluster management

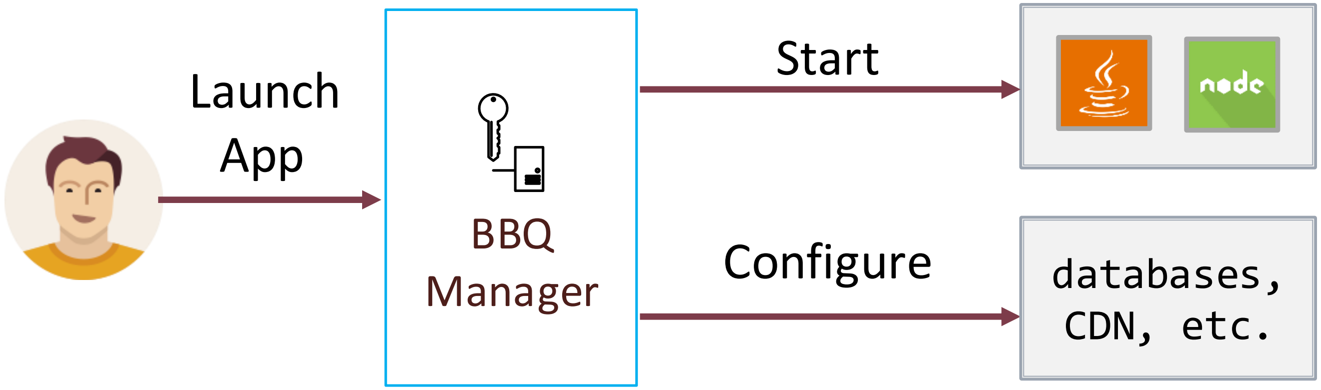
- Every tenant needs a private JVM and Node.JS process



Baqend Server

Customer‘s Business Logic

- Provisioning new instances needs to be fast & easy:



Launch App

BBQ Manager

Start

Configure

```
databases,
CDN, etc.
```

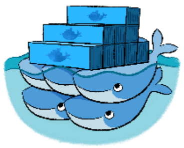
Problem: Many Technology Choices

Emerging Frameworks and Tools

- Cluster Managers & Orchestration Tools:

Google Kubernetes

Apache Mesos



Docker Swarm

Problem: Many Technology Choices

Emerging Frameworks and Tools

- Cluster Managers & Orchestration Tools:

Google Kubernetes

Apache Mesos

Docker Swarm

- Container Cloud Platforms:

Amazon Elastic Container Service

Tutum

Google Container Engine

Rancher

Problem: Many Technology Choices

Emerging Frameworks and Tools

- Cluster Managers & Orchestration Tools:

Google Ku

cker Swarm

and many more: Azure Container Service (Microsoft), Nomad (HashiCorp), Diego (Cloud Foundry), Fleet (CoreOs), ContainerShip, YARN (Hadoop), …

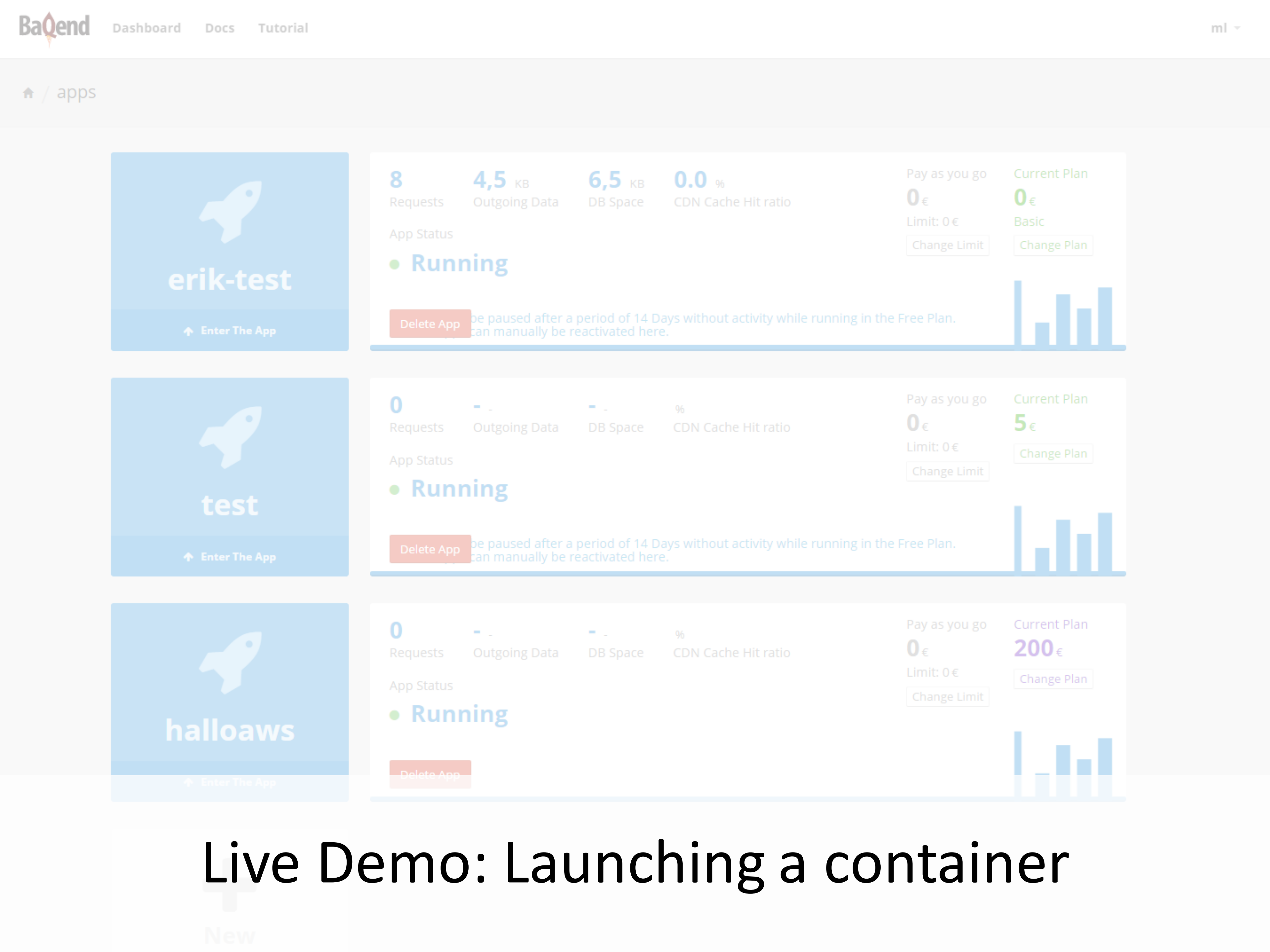
←Container

Amazon Elastic Container Service

Tutum

Google Container Engine

Rancher



BaQend Dashboard Docs Tutorial ml

apps

erik-test

Enter The App

8 Requests | 4,5 KB Outgoing Data | 6,5 KB DB Space | 0.0 % CDN Cache Hit ratio

App Status

● Running

Pay as you go 0 € Limit: 0 € Change Limit

Current Plan 0 € Basic Change Plan

Delete App be paused after a period of 14 Days without activity while running in the Free Plan. can manually be reactivated here.

test

Enter The App

0 Requests | - - Outgoing Data | - - DB Space | % CDN Cache Hit ratio

App Status

● Running

Pay as you go 0 € Limit: 0 € Change Limit

Current Plan 5 € Change Plan

Delete App be paused after a period of 14 Days without activity while running in the Free Plan. can manually be reactivated here.

halloaws

Enter The App

0 Requests | - - Outgoing Data | - - DB Space | % CDN Cache Hit ratio

App Status

● Running

Pay as you go 0 € Limit: 0 € Change Limit

Current Plan 200 € Change Plan

Delete App

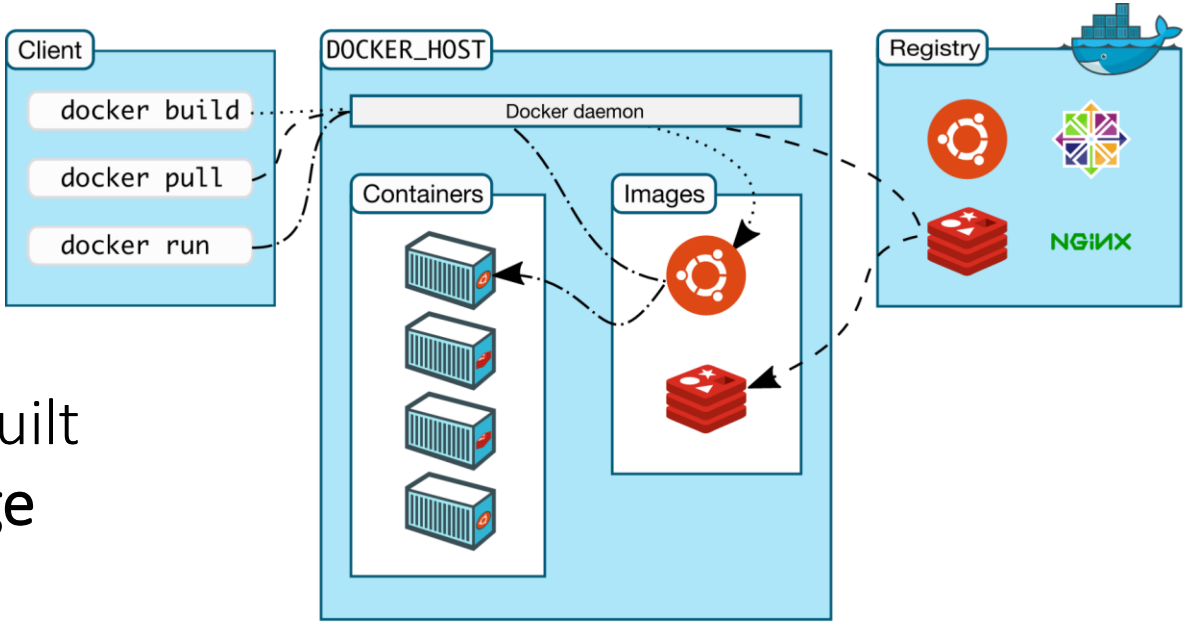
New

Live Demo: Launching a container

Docker Concepts

What is Docker?

- Docker typically isolates a **single application**
- An application is built into a **Docker image** (including the OS)



- The docker image can be hosted and transferred to different hosts (**Docker Registry**)
- The docker image can be executed as a new container on any machine that runs a **Docker daemon**
- **Updates** are handled by just stopping and starting a new container

Source: https://docs.docker.com/engine/introduction/understanding-docker/

Docker Architecture

How to set up a Docker host

- Docker runs on all common **Linux** distributions
- Docker can be installed from Docker's own package repository
- The Docker daemon can be configured by editing /etc/default/docker
- The Docker daemon allows many useful configurations:
    - Inter-container communication
    - Docker remote REST API
    - Labeling
    - DNS configuration
    - IP forwarding (disables internet for containers)
    - SSL encryption for the Docker damon

The Dockerfile

How to build a Docker image

```
FROM ubuntu:latest

ENV DEBIAN_FRONTEND noninteractive

# java
RUN apt-get install -y software-properties-common && \
    add-apt-repository -y ppa:webupd8team/java && \
    apt-get update && \
    echo debconf shared/accepted-oracle-license-v1-1 select true \
        | debconf-set-selections && \
    apt-get install -y oracle-java8-installer

# extract and install packages
ADD baqend-package*.tgz /opt
ADD config.json /opt/baqend/

EXPOSE 8080

WORKDIR /opt/baqend/

ENTRYPOINT ["java", "-classpath", "/opt/baqend/lib/*", "info.orestes.Launcher"]
CMD ["--config", "config.json"]
```

How a Docker container works

Isolation, performance, light-weight

- **Filesystem**: by using multiple read-only file systems and mounting a read-write file system on top
- **Data volumes**: mount additional physical volumes into the container
- **CPU**: by CPU shares and core limitation
- **Memory**: by defining memory constraints
- **Network**: by using virtual networks
- **System privileges**: such as port binding, execution rights, inter process communication, etc.
- **Logging**: by using docker logging capabilities or external loggers (json, syslog, aws, etc...)

Docker Options

Imposing constraints on containers

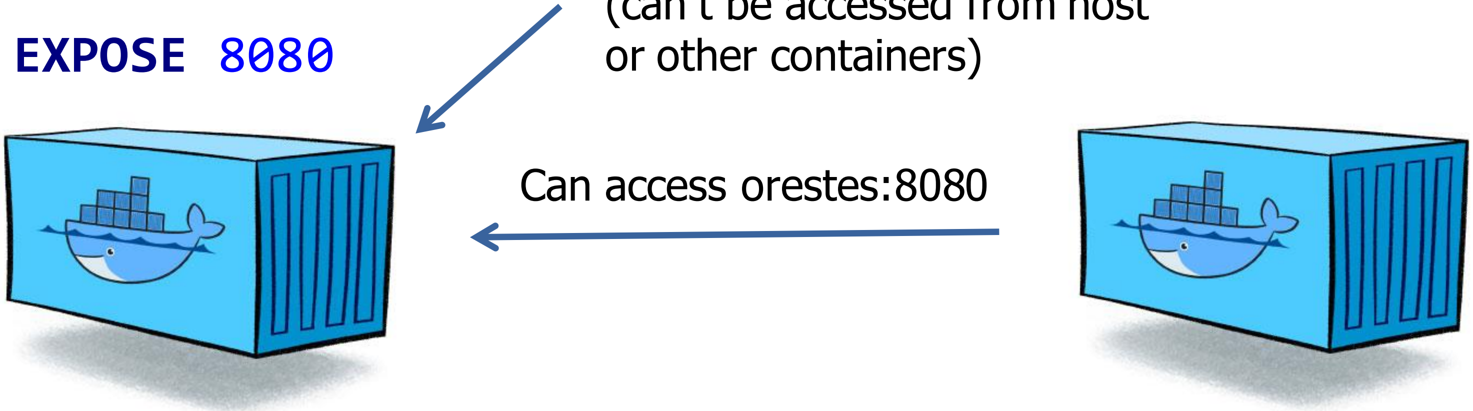
- Most constraints are set when the container is started

```
--add-host=[]               Add a custom host-to-IP mapping (host:ip)
--cpu-shares=0              CPU shares (relative weight)
--cpu-quota=0               Limit CPU CFS (Completely Fair Scheduler) quota
-e, --env=[]                Set environment variables
-l, --label=[]              Set metadata on the container (e.g., --label=key=value)
--link=[]                   Add link to another container
-m, --memory=""             Memory limit
--memory-swap=""            Total memory (memory + swap), '-1' to disable swap
--name=""                   Assign a name to the container
--net="bridge"              Connects a container to a network
                            'bridge': creates a new network stack on the docker bridge
                            'none': no networking for this container
                            'container:<name|id>': reuses another container network stack
                            'host': use the host network stack inside the container
                            'NETWORK': connects the container to user-created network
--oom-kill-disable=false    Whether to disable OOM Killer for the container or not
-p, --publish=[]            Publish a container's port(s) to the host
--read-only=false           Mount the container's root filesystem as read only
--restart="no"              Restart policy (no, on-failure[:max-retry], always)
-v, --volume=[]             Bind mount a volume
```

Docker Networking

Making containers talk to each other

- Docker containers can talk to each other by default
- Communication between containers can be restricted by the daemon option: `--icc=false`
- Docker containers can discover other linked containers by their names



```
EXPOSE 8080
```

Port 8080 not published, (can’t be accessed from host or other containers)

Can access orestes:8080

```
docker run --name="orestes" orestes
```

```
docker run --link="orestes" node
```

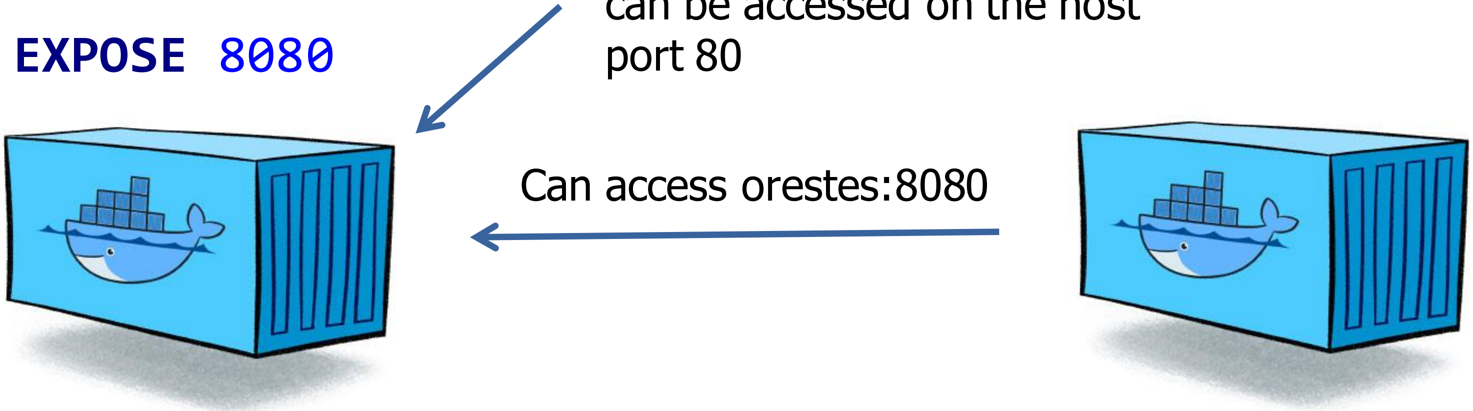
Docker Networking

Making containers talk to each other

- Docker containers can talk to each other by default
- Communication between containers can be restricted by the daemon option: `--icc=false`
- Docker containers can discover other linked containers by their names

```
EXPOSE 8080
```

Port 8080 is published and
can be accessed on the host
port 80



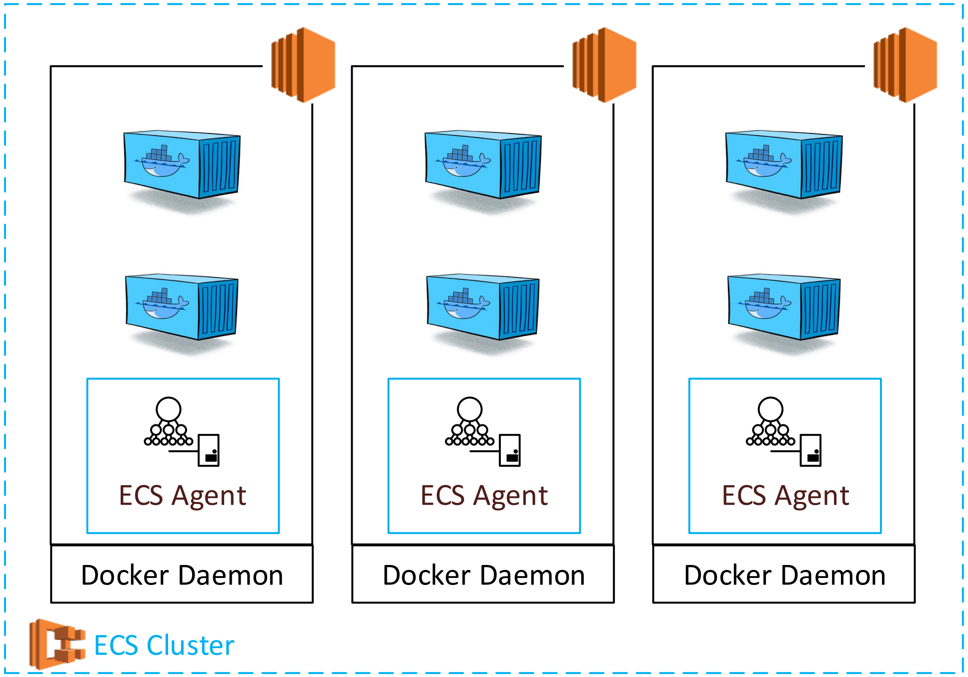
Can access orestes:8080

```
docker run --name="orestes"
    -p 0.0.0.0:80:8080 orestes
```

```
docker run --link="orestes" node
```

Elastic Container Service

How Amazon ECS works



- AWS provides ECS-**optimized AMIs** for simple deployment
- ECS manages EC2 instances by running an **ECS Agent** on each instance
- ECS can automatically deploy and scale new Docker containers specified by a **Task definition** across the ECS Cluster

ECS: Tasks and Services

Defining groups of containers

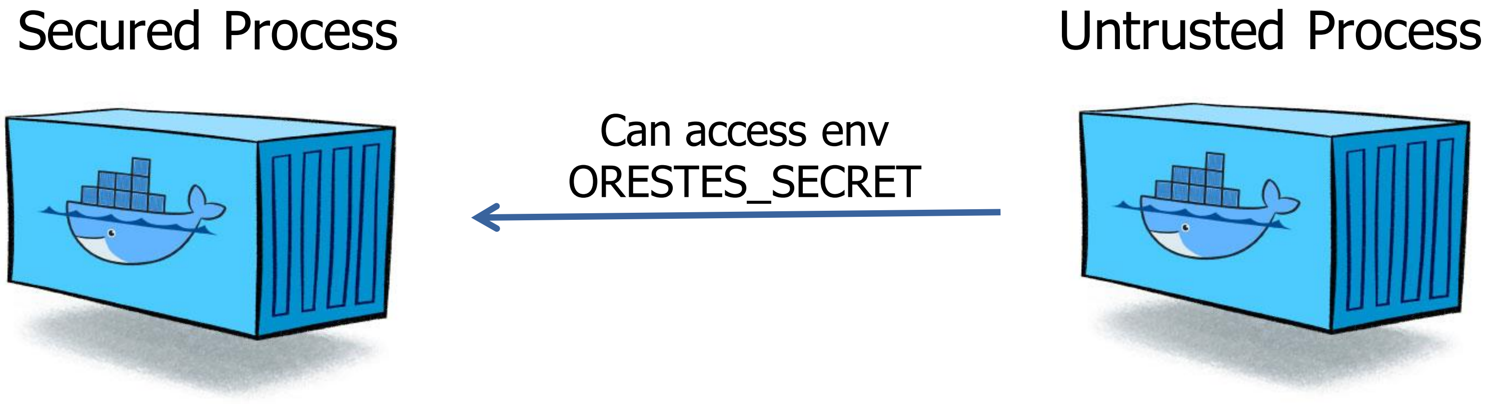
- ECS groups containers into Tasks and deploys them together
- A **Task definition** describes:
  - The Docker images
  - Resource requirements
  - Environment variables
  - Network links
  - Data Volumes
- ECS **Services** can be used to keep a specified number of Tasks running
- ECS can autoscale a Service when it is used with an ELB

Limitations that AWS fixed

Old Docker, Parameterization

- ECS has used an outdated version of docker, now it’s 1.9, yeah!
- Tasks can now be parametrized using commandline args

- Previously only environment variables could be passed while starting a Task
- Environment variables are exposed to linked containers, this can be a security issue!



Secured Process

Untrusted Process

Can access env ORESTES_SECRET

```
docker run --name="orestes"
    --env SECRET=7kekfjd9e
```

```
docker run --link="orestes" node
```

https://docs.docker.com/engine/userguide/networking/default_network/dockerlinks/#environment-variables

Current Limitation: Memory Constraints

Restricting RAM consumption

- ECS uses hard memory constraints (`run -m)` for Tasks to schedule container placement
- This allocates a **fixed amount of memory** on the EC2 instance and can't be exceeded by the process
- This is very ugly for shared, multi tenant applications:
  - Setting the constraint too low causes Docker to kill the process on memory peaks
  - Setting the value too high limits the number of containers that can be launched per EC2 instance
- Neither Docker's memory swapping nor unlimited memory usage is allowed by ECS

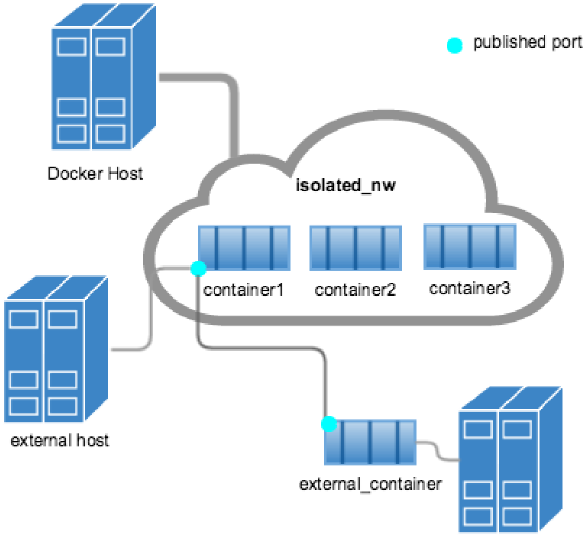
Current Limitation: Networking

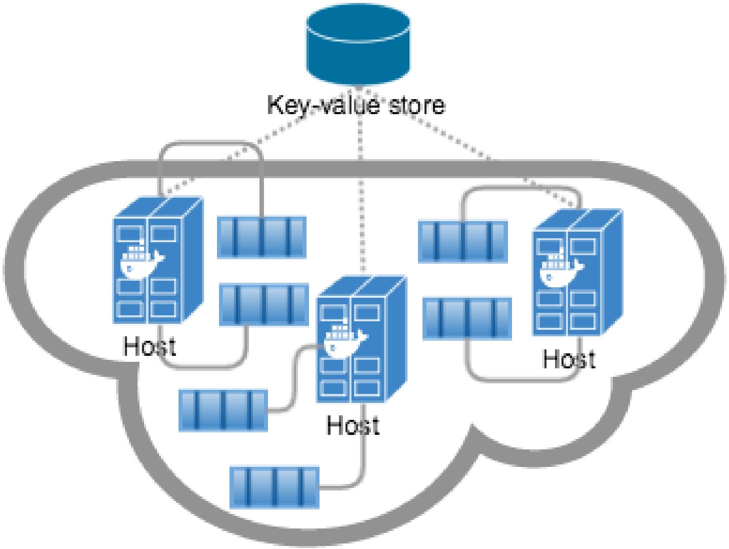
Docker's new network API not supported

- Docker has introduced a new network API, which allows to create custom virtual networks

- **Bridge Networks** connect groups of containers together and isolate them from other groups on the same host

- **Overlay Networks** use a key-value store to connect containers across different host machines





Source: https://docs.docker.com/engine/userguide/networking/dockernetworks/

Wrap-up: ECS

Pros and Cons

- Very **simple setup**, thanks to the optimized ECS AMI
- **Task** abstraction makes it really comfortable to start multiple containers together
- **Services** ensures that the desired count of tasks are always up and running
- **Automatically starts new EC2 instances** if no capacity is left for new containers
- Can be combined with an ELB for a **high availability** setup

- Many **Docker options** aren’t available
- Service Tasks can’t be **parametrized**
- **Running the same Services** for different tenants on the same EC2 instance is not possible
- Only the **legacy networking** is supported
- **New features** will always be delayed since they must first be implemented in ECS

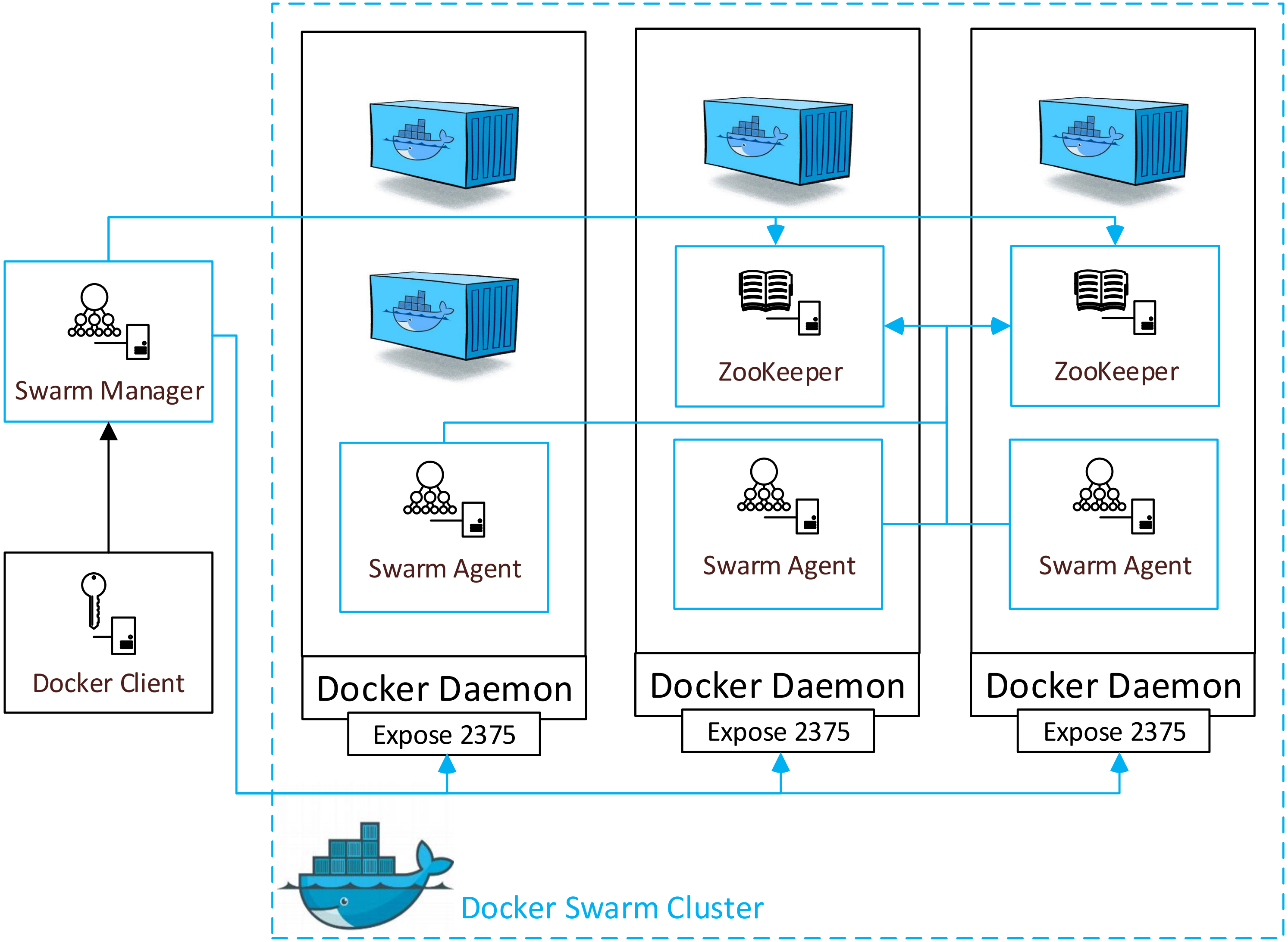
Docker Swarm

A replacement for ECS

- Docker Swarm is Docker's **native solution** for cluster management
- Docker Swarm uses a **discovery service** to manage the shared state of the cluster
- The following backends for discovery are supported:
  - Docker Hub (for development only)
  - Static file
  - etcd
  - consul
  - zookeeper
  - IP list or a range pattern of IPs

Swarm Architecture

Cluster management with Docker Swarm



Swarm is Docker

Fixing the shortcomings of ECS

- The Swarm manager acts as a **proxy** of the Docker Remote API
  - All Docker run options are available in Swarm, too
- Docker Swarm can be combined with **overlay networks**
  - Containers can connect to others by just using the containers name (**service discovery**)
  - Works across Docker hosts, availability zones and external hosts
- Containers can use any other service without defining them in a group (such as a Task)

Autoscaling in Swarm

Scale-out and scale-in

- Docker hosts can be added and removed to the Swarm Cluster silently
- Swarm provides an API to gather CPU usage and memory consumption of hosts or containers
- Swarm provides no concept to scale services within containers

High Availability in Swarm

Handling failures and outages

- **Labeled** Docker daemons can be used by the manager to run specific containers only on specific hosts
- Containers can be launched:
  - On the same host where other containers are running
  - In a specific availability zone
  - On hosts with special capabilities (RAM, CPU or SSD)
- The Docker daemon can **restart** failed containers using a restart policy `--restart="yes"`
- Containers will also be restarted if the docker host restarts
- Failed machines must be handled manually

Securing Swarm Hosts

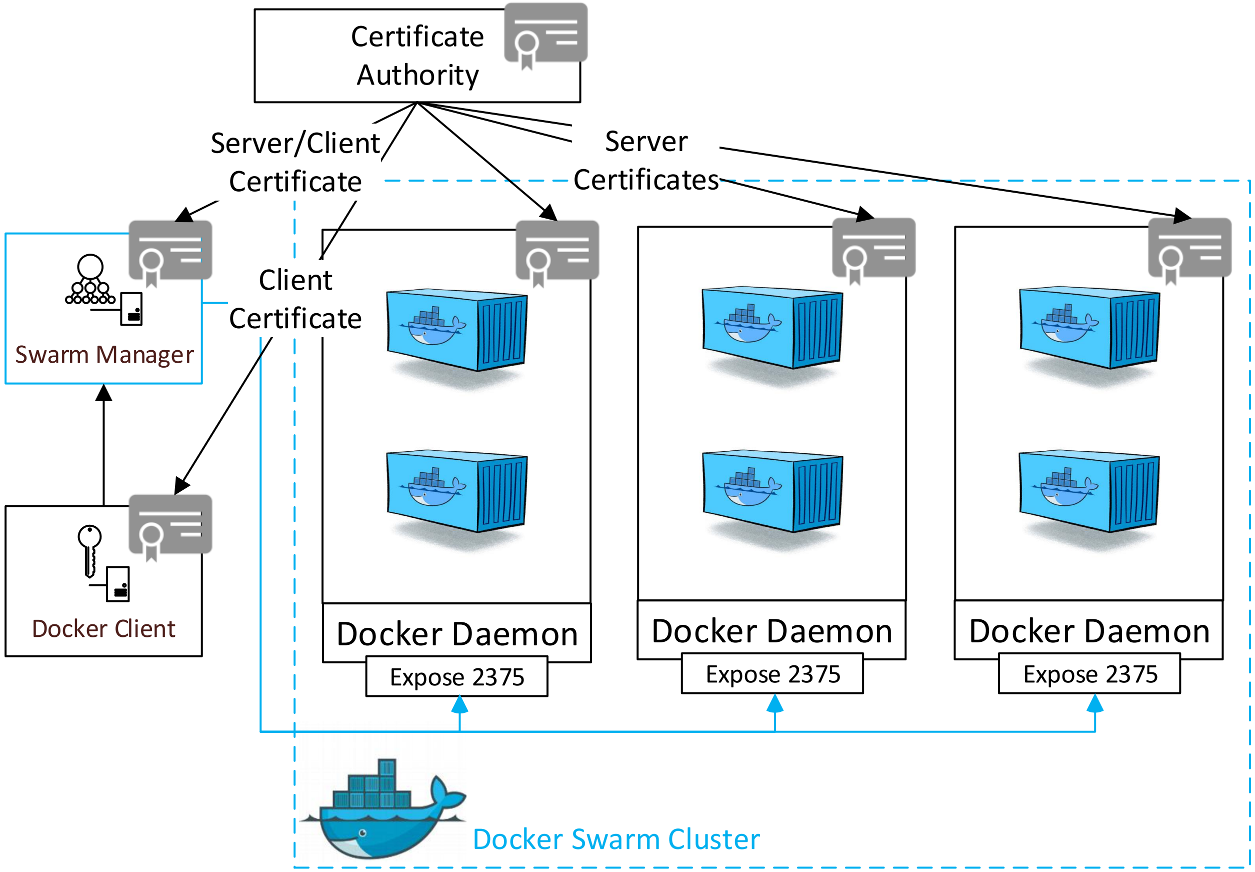
Security pitfalls

- Swarm requires that the Docker daemon is exposed via TCP
- In most setups this will be a security issue since you can easily get root permission on the Docker host
- Also containers can access the exposed API by default
- Therefore it is recommended to always secure the Docker daemons on each host with SSL
- Docker supports SSL client, server and both authentication mechanisms
- SSL server authentication is not very practical since it requires a signed certificate for each host

Securing Swarm Hosts

Security pitfalls

- Securing a Swarm cluster requires signed SSL certificates on all docker hosts, on the swarm manager and the docker client



Wrap-up: Docker Swarm

Pros and Cons

- **Swarm is Docker**, all Docker options are available
- **Labeling** Docker hosts, allows to deploy containers on specific hosts
- **Overlay Networks** allow containers to communicate across hosts
- **Service Discovery** across containers is made really simple

- **Complex setup** and many components are required for a complete setup
- No built-in way for **autoscaling** services
- Still many **bugs**
- The Docker Swarm **API integration** into Docker is not yet completed

Conclusions

ECS vs Swarm

- **Simple** Setup
- **Task** and **Service** definition makes it easy to deploy and update containers
- Detect **failures** and restart failed tasks within services
- **Integrated** into other AWS Services such as Elastic Load Balancers and Auto Scaling Groups

- **Complex** Setup
- Many **configuration** options for deploying containers
- Is compatible to the **Docker API**, allows to use all Docker clients
- Supports Docker's **network** API
- No **Vendor Lock-In**

Want to try Baqend?

- Download **Community Edition**
- Invited-Beta **Cloud Instance** *support@baqend.com*
- **Baqend Cloud** launching this February